R E P O R T R E S U M E S

ED 002 937 24

DEVELOPMENT AND ANALYSIS OF EXPERIMENTAL DESIGNS FOR RATINGS.

BY- STANLEY, JULIAN C. WILEY, DAVID E.
WISCONSIN UNIV., MADISON

REPORT NUMBER CRP-789 PUB DATE JUN 62
EDRS PRICE MF-$0.18 HC-$2.64 66P.

DESCRIPTORS- *TEACHER EVALUATION, EXPERIMENTAL PROGRAMS, DESIGN, *RATING SCALES, MEASUREMENT INSTRUMENTS, INTERPERSONAL PROBLEMS, *PSYCHOMETRICS, PERFORMANCE FACTORS, FACTOR ANALYSIS, *STATISTICAL ANALYSIS, *MATHEMATICAL APPLICATIONS, MADISON, WISCONSIN

EXPERIMENTAL PROCEDURES FOR RATING PERSONNEL WERE INVESTIGATED AS USED IN RESEARCH GENERALLY AND FOR USE IN THE EVALUATION OF TEACHERS. HISTORICAL PROBLEMS AND DEVELOPMENTS IN RATING TECHNIQUES WERE PRESENTED AND COMPARED WITH RANKING TECHNIQUES. STATISTICAL TECHNIQUES FOR ANALYSIS OF RATINGS WERE DETAILED AND RELATED TO VARIOUS TYPES OF RATING ERRORS. A MATHEMATICAL AND STATISTICAL APPARATUS USING A TWO-DIMENSIONAL ARRAY WAS DEPICTED TO ESTIMATE AND ISOLATE RATER BIASES, USING A TECHNIQUE WHICH INTERRELATED THE CONCEPTS OF AVERAGE INTERCORRELATION AND HYPOTHESIS TESTING VIA THE ANALYSIS OF VARIANCE PARTITIONING OF SOURCES OF VARIATION. THEOREMS WERE DEVELOPED EXTENDING THOSE CONCEPTS TO A N-DIMENSIONAL THEORY. THESE THEOREMS WERE APPLIED TO A FOUR-DIMENSIONAL CASE AND COMPLETE FORMULAS WERE SHOWN FOR CALCULATION OF AVERAGE VARIANCES AND COVARIANCES. FORMULAS FOR A PARTIAL NUMBER OF THESE VARIANCES AND COVARIANCES WERE ALSO GIVEN FOR A FIVE-DIMENSIONAL CASE. THE PROBLEMS INVOLVED IN RATING TEACHERS WERE DISCUSSED WITH PARTICULAR ATTENTION TO THE UNIQUE ASPECTS OF RATING TECHNIQUES. A STATISTICAL MODEL FOR TEACHER-RATING SCHEMES WAS DETAILED AND RELATED TO MODELS FORMULATED FOR OTHER APPLICATIONS. (JM)

ED 002937

Development and Analysis of Experimental Designs for Ratings

Julian C. Stanley and David E. Wiley

Laboratory of Experimental Design

Department of Educational Psychology

Education Building

University of Wisconsin

Madison 6, Wisconsin

June 1962

Project No. 789, Cooperative Research Branch, U. S. Office of Education,

Department of Health, Education, and Welfare

The background article for this project is:

Julian C. Stanley, Analysis of unreplicated three-way classifications, with applications to rater bias and trait independence. Psychometrika, June 1961, 26, 205-219.

The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare. Julian C. Stanley was the Principal Investigator. Research assistants who helped with one phase of the project or another were (in alphabetical order) Andrew B. Crider, David Freund, Mrs. Charlotte C. Holmes, and David E. Wiley. Mr. Wiley's special contributions to matrix formulation of the extension to k dimensions result in his being listed as coauthor, above.

This research will be reported at a meeting of the Psychometric Society and the APA Division of Evaluation and Measurement at the annual convention of the American Psychological Association in St. Louis, Missouri, early in September 1962. For details and an abstract see the July 1962 issue of the American Psychologist. Also, Mr. Wiley and Professor Stanley plan to prepare an article based upon this project for submission to Psychometrika.

The authors welcome your comments, criticisms, applications, and suggestions.

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

ABSTRACT of "Development and Analysis of Experimental Desgins for Ratings"*

Julian C. Stanley and David E. Wiley, University of Wisconsin

This report contains many new results in the field of ratings and several new applications of old results. We hope that it will be useful to those engaged in the evaluation of teachers and related pursuits.

The first chapter gives background to the difficult problem of ratings. It covers both intrinsic and extrinsic defects of ratings, defects which have hindered and will continue to hinder their interpretation. A brief summary of the concepts of rating design and some of its basic principles are given. Also the particular application to teacher rating is discussed. The mathematical conceptualization of rating matrices is explained in detail. This scheme of classification and conceptualization is detailed for several rating dimensions. The lengthiest section of this chapter includes an history of the problem of rating. It contains examples selected from the more than 1500 studies that were surveyed to show the increase in sophistication of the methodology and techniques of various investigators. Finally, an important distinction between the problems of rating and forced-choice ranking is made.

The second chapter is concerned with the history of the statistical development of the analysis of ratings prior to the compilation of this report. It outlines the original identification of the problem by Thorndike, who was the first to investigate and define the rating bias which he named "halo effect." The chapter also details some of the original methods of detecting and isolating these biases by factor analytic and other techniques. Next, the modern attack on the problem,

---

*U. S. Office of Education Cooperative Research Branch Project No. 789; Julian C. Stanley, Principal Investigator. June, 1962.

utilizing analysis of variance techniques, is described; it was first formulated fully by Guilford. Other viewpoints are discussed and related to this analysis of variance approach. In the last section of the chapter recent techniques, which are due to Stanley and which form the basis of this report, are described and their relationships with previous techniques are discussed.

In Chapter 3, the basic mathematical and statistical apparatus for the extension of Stanley's techniques beyond the case of two-dimensional rating schemes is given. These new techniques are valid for complete designs of any dimensionality. In the first part of the chapter, the basic definitions are given which are necessary for the development of the n-dimensional theory. Second, the basic theorems of this development are given. These theorems are then applied explicitly to the four dimensional case and the complete formulae for the calculation of needed average variances and covariances are given. Also the formulae for a partial number of these average variances and covariances in the five dimensional case are given.

The fourth chapter is concerned with the application of these techniques of analysis to the data from an actual experiment. The complete analysis of variance is given and all of the proper testing procedures are used. Also, the components of variance are computed. All of the average variances and average covariances are calculated using the techniques described in Chapter 3. The estimated average inter-correlations were calculated. Two of them are shown to be quite close to their actual values. The analysis of rating schemes is then compared with that of ranking schemes where various linear restrictions occur within the data.

The fifth chapter is called "Rating Teachers," and most of the chapter is concerned with the analysis of these particular types of rating schemes. The problems involved in and the particular unique aspects of this type of rating technique are discussed. The statistical

model appropriate to these rating schemes is given. This model is discussed in detail and related to models which have been formulated for other applications. In the light of the rating biases defined in earlier chapters and the restrictions imposed by the statistical model, proper adjustments for the elimination of rater biases are given. Finally the possibilities for the statistical treatment of designs which do not include all possible combinations of the levels of various factors are discussed. These "incomplete designs" will probably prove to be some of the most useful designs in rating situations.

The final chapter in this report is entitled "Cautionary Notes." It is basically concerned with two topics. The first is that of the relation of experimental design to the solution of psychometric problems. It is pointed out that the basic principles of experimental design need to be applied to psychometric research just as they are to more experimentally oriented research. In many cases lack of good design makes many seemingly straightforward inferences illegitimate. The last topic is concerned with the upper and lower numerical limits which an average correlation coefficient may attain. This change in limits from those of the usual correlation coefficient makes the average correlation more difficult to interpret.

In addition to these chapters, the report alsc has two appendices. The initial appendix contains the expectations of the mean squares under three mixed models and the expectations of the average variance and average covariances under one such model.

The second and last appendix contains the proof of the upper and lower bounds of the average correlation coefficient which were given in Chapter 6.

Chapter 1. Background of the Investigation

Rating techniques originated long ago, probably as an aspect of evaluating oneself and others, in which most persons engage rather constantly. Ordinary discourse abounds with statements such as "He is a good credit risk," "She learns arithmetic slowly," "I am an unusually friendly neighbor." Applied psychology and related disciplines have built upon this everyday use of rating procedures a body of technical lore, which is perhaps best set forth by Guilford in his chapter 11.[1]

Ratings prove to have a number of defects, in many instances much more serious than the faults that characterize test scores. Generally, they are situationally too unstable. Of course, it may make some difference whether one administers an objective vocabulary test in one place or another, or at one time of day or another, or in a selection (e.g. hiring) vs. classification context, but ratings tend to be influenced to a greater extent by such changes.

Also, the rating method itself can be crucial. How many rating-scale categories should be used? Free- or forced-choice? Anchored verbally or graphically?

If generalizations to other ratees and raters are desired, sampling becomes an important consideration.

We are fully cognizant of these threats to external validity.[2] Our concern in this project, however, is with internal validity--that is,

---

[1] J. P. Guilford, Psychometric methods (Rev. ed.). New York: McGraw-Hill, 1954.

[2] D. T. Campbell and J. C. Stanley, Experimental and quasi-experimental designs for research in teaching. N. L. Gage (Editor), Handbook of research on teaching. Chicago: Rand-McNally, 1962 (in press).

validity within a given rating "experiment." Independence of the traits used and various biases of raters toward ratees and traits deserve far more statistical attention than they have had until recently. Extending the design of rating experiments to some of the aspects usually treated in ad hoc fashion as external (situational conditioners, rating methods, and the like) would enlarge the applicability of our "internal" analyses.

Rating Teachers

Merit ratings of teachers are a special example of ratings in general. They have the usual sources of error, bias, and non-independence of traits, plus certain problems somewhat peculiar to schools. Typically, but not necessarily, a principal rates all his teachers once on each of a number of traits or characteristics. The frames of reference of the various principals usually differ rather markedly. Teachers differ greatly in age, teaching fields, sex, years of experience, abilities, and other factors, both within and between schools. These complicating complexities may make principals' ratings essentially unpredicatable. Ways to adjust the ratings to remove rater biases are available, if an extended, interlocking design requires each teacher to be rated by more than one rater (but not necessarily by all raters) in a systematic fashion, discussed later.

Matrices of ratings. If each of R raters (r = 1, 2, . . . , R) rates each of I individuals--that is, ratees--(i = 1, 2, . . . , I) S times (s = 1, 2, . . . , S) on each of T traits (t = 1, 2, . . . , T), an I x R x S x T four-way unreplicated classification of ratings is produced. This is an example of the basic data with which we are concerned. Ratings are illustrative, rather than exclusive. Any four-way classification of real numbers can be treated with the methods that we develop. For instance, if a group of students take two comparable forms of a certain test at the beginning of the school year

and the same two forms of that test at the end of the year, there are I students, two administration times, two forms, and Q questions per form, yielding in all 2 x 2 x I x Q item scores.[3]

The simplest rating situation of the above sort occurs when both I and R exceed 1 but S = T = 1; in other words, when every rater rates every ratee once on a single trait, such as sincerity. This yields a complete I x R matrix of ratings. The investigator may use this matrix to compute the R(R - 1)/2 coefficients of correlation among the R raters, thereby finding the extent of inter-rater agreement.[4] He may also want to extimate the variance component among ratees, intraclass r's,[5] and a coefficient of internal consistency[6] for the total scores over raters, $\sum_{r=1}^{R} X_{ir}$, where X_{ir} is the raw rating, X, received by the ith ratee from the rth rater.

The two-way I x R classification does not permit within-rater or between-trait comparisons, because each rater rates each ratee only once, on just one

---

[3] A situation somewhat like this is explored by J. C. Stanley in analysis of variance principles applied to the grading of essay tests. J. exp. Educ., 1962, 30 (in press).

[4] See K. A. Brownlee, Statistical theory and methodology in science and engineering. New York: Wiley, 1960, p. 254.

[5] See R. L. Ebel, Estimation of the reliability of ratings. Psychometrika, 1951, 16, 407-424, and E. E. Cureton, Note on the scaling of ratings or rankings when the numbers per subject are unequal, Psychometrika, 1952, 17, 397-399, and 1953, 18, ii.

[6] C. J. Hoyt, Relations of certain correlational to variance ratio estimates of test reliability. 12th Yrbk Natl Council on Measurements Used in Education, Part 1, 1955, pp. 50-55; Ebel, op. cit.; J. C. Stanley, K-R 20 as the stepped-up mean item intercorrelation, 14th Yrbk Natl Council on Measurement in Education, 1957, pp. 78-92.

trait. The three-way I x R x T classification makes between-trait comparisons possible, but not intra-rater intra-trait comparisons. Thus, for a full study of ratings, one needs at least a complete four-way classification. If more than one method of rating is to be studied, a fifth classification may be required. For the Campbell-Fiske validation matrix,[7] methods (m = 1, 2, . . . , M), such as test vs. questionnaire vs. direct observation, replace the r classification, requiring just four dimensions, but there are rating situations in which both R and M should exceed 1.

History of the Problem

Several years ago Stanley[8] investigated the two-way I x R case. Then, in February of 1959, while in Belgium as a Fulbright Act Research Scholar at the University of Louvain, he noted that the three-way I x R x T classification makes possible the isolation of three sources of covariance in place of the single one that the two-way classification permits: among traits within raters, among raters within traits, and across both raters and traits. The first source is related to the "relative halo effect" christened by Thorndike.[9] The second source is related to the independence of traits discussed by Willingham and Jones, McNemar, and Lord.[10] The third source is a "hetero-hetero" baseline against which the first two may be judged.

---

[7]D. T. Campbell and D. W. Fiske, Convergent and discriminant validation by the multitrait-multimethod matrix. Psychol. Bull., 1959, 56, 81-105.

[8]Stanley, op. cit., 1957.

[9]E. L. Thorndike, A constant error in psychological ratings. J. Appl. Psychol., 1920, 4, 25-29.

[10]W. W. Willingham and M. B. Jones, On the identification of halo through the analysis of variance, Educ. psychol. Measmt, 1958, 18, 403-407; McNemar, Attenuation and interaction.

Work on this three-way classification approach and its applications proceeded, along with extensions to higher dimensions, until mid-October of 1960, when the final form of a manuscript was accepted for publication.[11]

Literature survey. The principal investigator and a research assistant, Mr. David Freund, scrutinized some 1500 studies involving ratings in an attempt to locate data matrices of three or more dimensions and ascertain what experimental and statistical procedures, if any, were employed to minimize relative halo effect and to maximize independence of traits. Few studies met the more-than-two-dimensions requirement. That is, few investigators had secured ratings of more than one ratee by more than one rater on more than one trait.

The most common design, found in a large number of attempts to predict teacher "success," involved having each school principal rate or rank his own teachers on one or more traits, thereby making R, the number of raters for a particular group of teachers, 1. This design rules out the possibility of ascertaining the main effect of r and the two interactions i x r and r x t. It also makes the i x t interaction untestable. We searched in vain for studies reporting clever interlocking procedures that result in less than full sets of data but allow rater biases (r, i x r, and r x t) and trait independence (i x t) to be evaluated. Cornell's[12] observers, who visited each class in pairs several times and rated independently of each other, are probably a desirable alternative to the perhaps intrinsic invalidity of principals' ratings.

---

[11] J. C. Stanley, Analysis of unreplicated three-way classifications, with applications to rater bias and trait independence. Psychometrika, 1961, 26, 205-219.

[12] F. G. Cornell, C. M. Lindvall, and J. L. Saupe, An exploratory measurement of individualities of schools and classrooms. Univ. of Ill. Bull. No. 75, 1953.

The oldest relevant study that we found appeared in 1915.[13] This was an interesting attempt to determine how variable the ratings of teachers by a number of principals are because of different rating rationales of these principals. One hundred and ten principals each rated ten hypothetical teachers represented by verbal descriptions, on each of two characteristics (instruction and discipline), yielding a 110 x 10 x 2 matrix of ratings on a 5-point scale (A = excellent, B+ = very good, B = good, C = fair, and D = deficient) that might be treated quantitatively as 4, 3, 2, 1, 0. Unfortunately, Shiels does not preserve distinctions among the 110 principals, except to summarize the number and percentage who gave a rating of A to Case 1 on discipline, etc. The range of ratings of the same case on the same characteristic by the 110 principals was uncomfortably large. To Shiels "the data presented show the necessity of some more detailed analysis of the teacher's function" (p. 745). Small wonder, when "Under 'instruction' are included 'teaching ability,' 'scholarship' and 'effort.' The term 'discipline' embraces 'personality,' 'control of class' and 'self-control'" (p. 752). Less composite characteristics seem required.

Most of the attention to rating of traits came after World War I, however, in the burst of applied psychological activity following experiences with the Army Alpha and Beta and the Woodworth Personal Data Sheet.[14] By 1925 this resulted in an article entitled "An experimental study of the rating scale technique"[15] that reported a large study involving three different types of

---

[13] A. Shiels, The rating of teachers in the New York City public schools. Sch. & Soc., 1915, 2, 752-754.

[14] S. I. Franz, Handbook of mental examination methods. New York: Macmilla 1919.

[15] By Sarah E. Marsh and F. A. C. Perrin, J. abnorm. soc. Psychol., 1925, 19, 383-399.

rating scales. With small modifications of the experimental procedure, these authors could have had a four- or five-dimensional matrix of ratings, suitable for detailed analysis in the way that we propose later in this report. In keeping with the methodology and psychology of that period, though, Marsh and Perrin were more interested in intercorrelations of their trait ratings with other evidences of the traits (concurrent validity) than in the "construct" validity concept that has developed from 1954 onward. They were aware of "relative halo effect," though apparently not of the other types of rater bias that might have been isolated from their data if appropriate techniques had been available then. Nevertheless, for 1925 this was an exceptionally well conducted and efficiently analyzed study.

Nanninga[16] studied both inter- and intra-rater variation, the latter over a three-week period. Using a five-category scale, five teachers rated 30 pupils twice on 15 traits. "Only 20 pupils out of the 30 were rated by all five teachers on the 15 traits since each teacher did not have all the pupils" (p. 115). Nanninga's ratings for these 20 pupils would be amenable to the four-dimensional analysis that we describe later. He teased out enough information about ratings of a given trait both within and between teachers to conclude that "teachers are not consistent with their own ratings on a number of the traits" (p. 116) and the "The vast difference in agreement among teachers suggests that individual prejudices play a part" (p. 118). One strong prejudice may have been against Nanninga, who apparently coerced these teachers via their principals to devote 1-3 hours to the first rating endeavor, and then several weeks later asked them to do the same thing again, having merely rearranged the order of the traits on the rating sheets. It would be interesting to know how perfunctorily they made the second ratings.

---

[16] S. P. Nanninga, A critical study of rating traits. Educ. Admin. & Superv., 1926, 12, 114-119.

It is of course not surprising that considerations of randomization and variance analysis are lacking from the Shiels, Marsh & Perrin, and Nanninga studies, for the pioneering statistical research of Fisher did not reach book form until 1925,[17] nor did it become influential in education and psychology until the 1940's.[18]

Adams[19] had each of 10 raters rank themselves and the 9 other raters twice on each of 63 traits. He then studied self-rankings vs. rankings by others. His ranking procedure eliminates "level" variation among raters, because the sum of the consecutive untied ranks from 1 through n assigned by a given rater is always $n(n + 1)/2$. Likewise, it prohibits any t or r x t variation, because a given rater cannot rate one trait higher overall than another trait. Ratees are not so restrained, however. They may exhibit great differences among themselves and interact with raters and with traits. Because the sum of ranks and the sum of squared ranks in uniquely determined

---

[17] R. A. Fisher, Statistical methods for research workers. Edinburgh: Oliver & Boyd, 1925 (1st edition).

[18] Perhaps the two books most responsible for popularizing the analysis of variance and covariance in these two fields were E. F. Lindquist's Statistical analysis in educational reserach (Boston: Houghton Mifflin, 1940) and Quinn McNemar's Psychological statistics (New York: Wiley, first edition 1949). Even yet, however, randomization is seldom used in rating studies, though the statistical analyses employed are typically predicated on it. This may invalidate customary analyses, as Oscar Kempthorne and his students at the Iowa State University are showing in their elucidation of what they term "structures."

[19] H. F. Adams, The good judge of personality. J. abnorm. soc. Psychol., 1927, 22, 172-181.

by the ranking procedure, the total variance is not free to vary from study to study of a given size, so the highest-order interaction (i x r x s x t) sum of squares is merely a simple linear function of the other SS's. This makes certain significance tests uninterpretable, because the ratios of mean squares used do not have independent numerator and denominator and hence do not follow the F distribution even if the null hypothesis is true.

We make the above technical points merely to show that, while formally similar, rating and ranking require rather different methods of analysis. Much more powerful procedures than Adams had at his disposal now exist,, but not all statistical problems relative to forced-choice techniques such as ranking have yet been solved.[20]

Garrison and Howell[21] had three teachers rate each of 62 pupils on 19 character traits, which they divided into two classifications: mental and activity traits, and social and temperamental traits. Like Marsh and Perrin,[22] they were interested in the agreement of ratings with objective data. Also, they used ratings by pupils of other pupils as criteria. Again, the complete Campbell-Fiske[23] type of analysis is lacking, so that not all information in the data is utilized. Typically for that time, analysis is via correlation coefficients and deviations from norms.

---

[20]For discussion of this point, see W. C. Krumbein and J. Tukey, Multivariate analysis of mineralogic, lithographic, and chemical composition of rock bodies. J. sedimentary Petrology, 1956, 26, 322-337.

[21]K. C. Garrison and Sue Craft Howell, The Relationship between character trait ratings and certain mental abilities. J. appl. Psychol., 1931, 15, 378-389.

[22]Marsh and Perrin, op. cit.

[23]Campbell and Fiske, op. cit.

The sixth, and final, historical reference that we shall cite here to show that even many years ago there existed designs permitting the type of analysis we have devised was reported by Wolf and Murray[24] in 1937. This is an excellent segment from an even larger study. Methodological sophistication is evident throughout the paper. Five judges observed 28 college men one at a time for 45 minutes each and then independently marked each man on each of 40 variables, using a five-point scale.. This is the standard three-way classification, for which detailed analytical techniques have been offered recently by Campbell and Fiske[25] and by Stanley.[26] Wolf and Murray did not use such procedures then, of course.

A summary of methods of investigating the effectiveness of teachers appeared recently.[27]

Published studies from 1915 to the present day reveal abundantly how needed are the average covariance techniques that we develop.

---

[24]R. Wolf and H. A. Murray, An experiment in judging personalities. J. Psychol., 1937, 3, 345-365.

[25]Campbell and Fiske, op. cit.

[26]Stanley, 1961, op. cit.

[27]A. S. Barr and others, Wisconsin studies of the measurement and prediction of teacher effectiveness, a summary of investigations. J. exp. Educ., 1961, 30 (entire September issue).

Chapter 2. Statistical Development

The treatment of the expression for average inter-covariances in terms of the mean-squares in an n-way analysis of variance arises logically from a sequence of articles and books published since Thorndike's initial recognition of the "halo effect" in 1920.[1] Chi seems to have been the first to attempt a detailed statistical analysis or evaluation of halo effect in rating scales.[2]

After these publications there was a lengthy interval before Guilford[3] treated rater biases and effects in terms of the analysis of variance. Then came Johnson and Vidulich's[4] utilization of this scheme for the experimental manipulation of one of these biases.

Out of sequence, but very pertinent to this analysis, was Gulliksen's[5] first formulation in the unreplicated 2-way case, of the basic identities of interest in the present study. From this basic formulation, Willingham and Jones[6] recognized the utility of Gulliksen's work for finer statistical dis-

---

[1] E. L. Thorndike, A constant error in psychological ratings. J. appl. Psychol., 1920, 4, 25-29.

[2] P.-L. Chi, Statistical analysis of personality rating. J. exp. Educ., 1937, 5, 229-245.

[3] J. P. Guilford, Psychometric methods (2nd ed.). New York: McGraw-Hill, 1954.

[4] D. M. Johnson and R. N. Vidulich, Experimental manipulation of the halo effect. J. appl. Psychol., 1956, 40, 130-134.

[5] H. Gulliksen, Theory of mental tests. New York: Wiley, 1950, Pp. 54-56.

[6] W. W. Willingham and M. B. Jones, On the identification of halo through the analysis of variance. Educ. psychol. Measmt, 1958, 18, 403-407.

crimination and manipulation of basic rating biases and effects.

Finally, Stanley[7] showed how all of these concepts and statistical tools could be brought together and unified into a powerful method of analysis.

One way to measure halo effect is to use the difference between two intercorrelations. Chi[8] was perhaps the first to measure halo in this way. He conceptualized halo as an effect due to an individual rater's being generally biased toward certain individual ratees. This definition implies that the intertrait correlations for a single rater should be high. When averaged over all raters, this average intra-rater intertrait correlation should remain high.

The problem remains, however, relative to what is the average intra-rater intertrait correlation high? Chi felt that it should be higher than the average intercorrelation across both raters and traits. This intercorrelation is the correlation between a given rater on a given trait with another rater on another trait. When these are averaged, we have the average intercorrelation across raters and traits. Chi used the difference between the average intra-rater intertrait r and the average inter-rater intertrait r to measure the relative halo effect. Chi was then able to estimate the amount of variability due to halo by means of a factor analysis.

J. P. Guilford put the problem of rating in perspective when in the 1954 edition of Psychometric methods[9] he identified and classified the following types of rating errors:

---

[7]J. C. Stanley, Analysis of unreplicated three-way classifications, with applications to rater bias and trait independence. Psychometrika, 1961, 26, 205-219.

[8]P.-L. Chi, Statistical analysis of personality rating. J. exp. Educ., 1937, 5, 229-245.

[9]New York: McGraw-Hill, 1954. Pp. 278-281.

1) Leniency error--this consists of constant bias of raters over ratees and traits. One of the causes of this type of error is indicated by its name.

2) Halo effect--this is a bias of an individual rater toward an individual ratee. The effect of this error is to increase the intercorrelations between the traits.

3) Central tendency error--this describes the tendency of most raters to give judgments toward the center of the scale and not to use the extreme categories.

4) Logical error--this describes the failure of ratees to make a distinction between some of the traits, for example, the failure to distinguish between honesty and truthfulness. This error also increases the intercorrelation between the traits.

5) Contrast error--this is an error which denotes the tendency of a rater to rate a ratee low in a trait which he considers himself high on. This error would then cause a rater to react differentially toward different traits.

Guilford proposed that some of these errors can be tested for significance and perhaps estimated by means of a factorial analysis of variance. He proposed that the following main effect and interactions would measure some of the errors given above:

The rater main effect (r) is a measure of the differential leniency errors of various raters. This is true since it measures differences in the means of the raters.

The rater-ratee interaction (r x i) is a measure of halo effect. This is true since the interaction measures the fact that some raters react differently toward some ratees.

The rater-trait interaction (r x t) is a measure of contrast and

similar errors. This is because these errors result in raters' reacting differentially to the individual traits.

With these effects identified as measuring the various errors indicated, Guilford proceeds to observe that they may be tested and identified through a standard analysis of variance.

Once an analysis is done and an effect identified as present, Guilford recommends methods of eliminating it. For instance if the i x r (halo) effect is significant, he recommends the following adjustment of the $\bar{X}_{ir.}$ means:

$$\bar{X}'_{ir.} = \bar{X}_{i..} + \bar{X}_{.r.} - \bar{X}_{...} \quad .$$

Following in a similar manner and assuming that all the errors (r, i x r, and r x t) are present, he gives a procedure equivalent to the following formula for total correction:

$$X'_{irt} = X_{irt} - \bar{X}_{ir.} - \bar{X}_{.rt} + \bar{X}_{i..} + \bar{X}_{.r.} + \bar{X}_{..t} - \bar{X}_{...} \quad .$$

The "reverse" biases caused by use of this formula will be covered later in this report.

So it can be seen that the significant idea which Guilford contributed to the analysis of ratings was the introduction of the analysis of variance to isolate and test some of the various errors in rating scales.

In 1956, Johnson and Vidulich[10] proposed a method of reducing halo effect and measuring the effectiveness of the process by use of Guilford's analysis of variance approach. They let two groups of raters rate a group of prominent people on various traits. One group was forced to rate the I ratees in a way designed to maximize halo effect and the other in a way designed to minimize it.

In order to maximize halo effect, Johnson and Vidulich inverted Guilford's suggestion that it might be minimized by presenting each trait separately with

---

[10]Johnson and R. N. Vidulich, op. cit.

all ratees to be rated at the time on the trait; they presented all traits on a single sheet with a single ratee. Also, they tried to minimize halo effect for another group of raters by following Guilford's suggestion and presenting each trait separately with all ratees to be rated. Separated analyses of varianc were done on each.

The results are given in the following table:

Minimization	Maximization	Interaction
$P < .05$	Not significant	r x t
Not significant	$P < .01$	i x r

They concluded that the minimization procedure was effective in reducing halo effect, but that it resulted in an increase in differential reaction of raters with traits.

When Johnson[11] reanalyzed the data with more valid statistical techniques he found that the (i x r) and (t x r) interactions were both significant using both techniques:

Minimization	Maximization	Interaction
$P < .01$	$P < .05$	r x t
$P < .01$	$P < .01$	i x r

The F-ratios changed in the directions indicated on the original analysis but the two conditions were not significantly different.

In his Theory of mental tests, Gulliksen[12] derived the expression for writing mean squares in the analysis of variance in terms of average variances

---

[11] D. M. Johnson, Reanalysis of experimental halo effects. Mimeographed, 1961.

[12] Gulliksen, op. cit.

and covariances. On page 54 he writes:

$$\frac{I^2}{(N-1)(K-1)} = \overline{(s^2_{x_g})} - \overline{r_{gh}s_g s_h}$$ (where I^2 represents the interaction sum of squares),

which when expressed in our terminology is written as:

$$MS_{(i \times g)} = \overline{s_g^2} - \overline{\mathrm{cov}(X_g, X_{g'})} .$$

This expression, it must be noted, is valid only for a two-dimensional array, where the scores are of the form X_{ig}. In a similar manner Gulliksen derived the following representation for the sum of squares among means of the individuals:

$$\frac{P^2}{N-1} = \overline{(s_g^2)} + (K-1)\overline{r_{gh}s_g s_h}$$ (where P^2 represents the sum of squares among the individual means),

In our notation this may be expressed as:

$$MS_i = \overline{s_g^2} + (K-1)\overline{\mathrm{cov}(X_g, X_{g'})}$$, where g = 1, 2, . . . , K.

These two formulae form the basis of the analysis which appears below.

Willingham and Jones[13] proposed a further clarification of the errors identifiable through the analysis of variance. They also reformulated Guilford's analysis into an overall analysis of variance. Their classification and clarification are as follows:

1) Rater main effect, r, measures leniency error;

2) Ratee main effect, i, measures the variance between ratees plus "absolute halo," an effect common to raters (an overall rater bias);

3) Trait main effect, t, measures the variance between traits plus an overall rater bias of rating some traits high and others low;

4) Rater-trait interaction, r x t, is the "contrast" tendency for different raters to rate traits differentially;

---

[13] Op. cit.

5) Rater-ratee interaction, i x r, 'relative halo,' shows the tendency for different raters to rate ratees differentially.

Willingham and Jones were also interested in the third type of interaction, the ratee-trait interaction. This i x t interaction they say "represents the extent to which the raters as a group make trait-rater discriminations." They recall that a common measure of halo is the average intertrait correlation, $\bar{r}_{tt'}$. They then make use of Gulliksen's formula, which they state in a form equivalent to $MS_{(i \times t)} = \overline{s_t^2} - \overline{r_{tt'} s_t s_{t'}}$ $(t \neq t')$, where the i x t interaction is taken over the two-dimensional array of $\bar{X}_{i.t}$ means. The conclusion to which they come is that the i x t interaction is an inverse measure of the type of halo that is normally measured by the average intertrait correlation. This type of halo they call 'composite halo.' This is in our framework, a measure of trait independence; see Stanley.[14]

Following the line of Willingham & Jones and Gulliksen, we may demonstrate the relationships between the mean squares and the average covariances again in the following manner:

$$MS_i = \overline{s_t^2} + (K - 1)\overline{cov(X_t, X_{t'})} \quad (t = 1, 2, \ldots, K);$$

$$MS_{(i \times t)} = \overline{s_t^2} - \overline{cov(X_t, X_{t'})} .$$

This gives rise to the matrix equation:

$$\begin{pmatrix} 1 & k - 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \overline{s_t^2} \\ \overline{cov(X_t, X_{t'})} \end{pmatrix} = \begin{pmatrix} MS_i \\ MS_{(i \times t)} \end{pmatrix} .$$

Since this expression is valid for any two-dimensional array of real numbers, the question arises as to whether this result can be extended to three, four, or higher dimensional arrays. Obviously this would be very useful in the three dimensional case; Willingham and Jones were limited in their por-

---

[14] Op. cit.

trayal of this case by the fact that Gulliksen's formulation was limited to two factors. They used only the mean values $\bar{X}_{i.t}$ in a two-dimensional array. One of the main goals of our report is to make this extension to three dimensions and higher.

After a review of these articles, we can see that there are three main statistical concepts utilized in them which are pertinent to the estimation and isolation of rater biases. One of these, average intercorrelation, is essentially estimational. Another is oriented toward hypothesis testing via the analysis of variance partitioning of sources of variation. The third concept, and the most important one for the purposes of this report, is that which connects and makes clear the interrelations between the first two.

It can be seen than that the primary result of this report is to unify the correlational concepts of Chi and others with the experimentally oriented statistical concepts of Guilford and others by making use of the basic relationships derived by Gulliksen. It is hoped that this newly derived relationship between the two approaches will bring about more research and detailed analysis in this area.

Chapter 3. Extension beyond Two Dimensions

In this chapter we proceed directly with the mathematical development that leads to a recursion formula in matrix notation for finding the mean variance and the mean covariances for a complete unreplicated factorial design with any number of factors 1, 2, 3, 4, ... We then provide all formulas for the four-factor analysis and the formulas for covariances within individuals in the five-way design. All formulas for the three-way case were provided by Stanley.[1]

[1] J. C. Stanley, Analysis of unreplicated three-way classifications, with applications to rater bias and trait independence. Psychometrika, 1961, 26, 205-219. Formulas 1-24.

I. Notation for variances and covariances

Each integer h, where $0 \leq h \leq 2^n$ (n is the number of added factors and 2^n is the number of variances and covariances) can be expressed in terms of binary digits. Let $h = \sum_{g=1}^{n} k_g 2^{g-1}$, where each $k_g = 0$ or 1, depending on conditions specified below.

Now in each average variance and covariance some of the variables in the subscripts are primed and some are not. If we define $k_g = 0$ when the gth variable (added factor) is not primed and $k_g = 1$ when the gth variable is primed, then we are in a position to make the following definition:

Definition:

$$c_h^{(n)} = \frac{\sum_{i_n=1}^{N_n(N_n^{k_n} - k_n)} \cdots \sum_{i_1=1}^{N_1(N_1^{k_1} - k_1)} \sum_{i_1=1}^{N_0} (X_{i_0 i_1 \cdots i_n} - \bar{X}_{\cdot i_1 \cdots i_n})(X_{i_0 i_1}(k_1) \cdots i_n(k_n) - \bar{X}_{\cdot i_1}(k_1) \cdots i}{(N_0 - 1)\prod_{g=1}^{n} N_g(N_g^{k_g} - k_g)}$$

where $i_g^{(k_g)} = i_g$ if $k_g = 0$,

$= i_g'$ if $k_g = 1$.

We may note that the expression $N_g(N_g^{k_g} - k_g) = N_g$ if $k_g = 0$, and

$= N_g(N_g - 1)$ if $k_g = 1$.

Also, i_g represents the subscript corresponding to the gth added factor.

Example: $c_0^{(n)} = \dfrac{\sum_{i_n=1}^{N_n} \cdots \sum_{i_1=1}^{N_1} \sum_{i_0=1}^{N_0} (X_{i_0 \cdots i_n} - \bar{X}_{\cdot i_1 \cdots i_n})^2}{(N_0 - 1)\prod_{g=1}^{n} N_g} = \overline{s^2_{i_1 \cdots i_n}}$;

$$c_3^{(4)} = \frac{\sum_{i_4=1}^{N_4}\sum_{i_3=1}^{N_3}\sum_{i_2=1}^{N_2(N_2-1)}\sum_{i_1=1}^{N_1(N_1-1)}\sum_{i_0=1}^{N_0}(X_{i_0i_1i_2i_3i_4} - \bar{X}_{.i_1i_2i_3i_4})(X_{i_0i_1'i_2'i_3i_4} - \bar{X}_{.i_1'i_2'i_3i_4})}{(N_0 - 1)N_1(N_1 - 1)N_2(N_2 - 1)N_3N_4} .$$

since $3 = 1(2^1 - 1) + 1(2^2 - 1) + 0(2^3 - 1) + 0(2^4 - 1)$. Therefore $k_1 = k_2 = 1$, and $k_3 = k_4 = 0$.

Finally, from the definition, $c_3^{(4)} = \overline{\text{cov}(X_{i_1i_2i_3i_4}, X_{i_1'i_2'i_3i_3})}$.

Example:

From this definition we can see the following:

$\overline{\text{cov}(X_{rtsm}, X_{r't'sm})} = c_{0011}^{(4)} = c_3^{(4)}$;

$\overline{\text{cov}(X_{rstm}, X_{r'ts'm})} = c_{0101}^{(4)} = c_5^{(4)}$;

$\overline{\text{cov}(X_{rstm}, X_{r't's'm'})} = c_{1111}^{(4)} = c_{15}^{(4)}$;

$s_{rstq}^2 = c_{0000}^{(4)} = c_0^{(4)}$.

Definition: $c_n = \begin{vmatrix} c_0^{(n)} \\ \vdots \\ c_{2^n-1}^{(n)} \end{vmatrix} = (c_h^{(n)})$, $h = 0, \ldots, 2^n - 1$.

II. Notation for Mean Squares

In a manner similar to above:

$m_h^{(n)} = MS_{(h)}$, where the mean square is from an n-dimensional table, $h = 0, \ldots, 2^n - 1$, and each

k_g in h is 0 or 1 depending on whether or not the gth factor is a part of the particular interaction.

$$MS_{(i \times t)} = m^{(5)}_{0010} = m^{(5)}_2$$

$$MS_i = m^{(5)}_{0000} = m^{(5)}_0$$

$$MS_{(i \times t \times s)} = m^{(5)}_{0110} = m^{(5)}_6$$

$$MS_{(i \times r \times t \times s \times q)} = m^{(5)}_{1111} = m^{(5)}_{15}$$

<u>Definition</u>: $m_h = \begin{pmatrix} m^{(n)}_0 \\ \vdots \\ m^{(n)}_{2^n - 1} \end{pmatrix} = \left(m^{(n)}_h \right)'$.

III. Notation for Transformation Matrix

$A_n = (a^{(n)}_{jh})$, $j,h = 0, \ldots, 2^n - 1$.

$$A_n = \begin{pmatrix} A_{n-1} & (N_n - 1)A_{n-1} \\ A_{n-1} & -A_{n-1} \end{pmatrix}$$

where N_n is the number of levels of the <u>n</u>th added factor. $A_0 = 1$. N_0 = the number of levels of the original factor.

<u>Example</u>:

$$A_2 = \begin{pmatrix} A_1 & (T-1)A_1 \\ A_1 & -A_1 \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} A_0 & (R-1)A_0 \\ A_0 & -A_0 \end{pmatrix} & (T-1)\begin{pmatrix} A_0 & (R-1)A_0 \\ A_0 & -A_0 \end{pmatrix} \\ \begin{pmatrix} A_0 & (R-1)A_0 \\ A_0 & -A_0 \end{pmatrix} & -\begin{pmatrix} A_0 & (R-1)A_0 \\ A_0 & -A_0 \end{pmatrix} \end{pmatrix}$$

Therefore:

$$A_2 = \begin{pmatrix} A_0 & (R-1)A_0 & (T-1)A_0 & (T-1)(R-1)A_0 \\ A_0 & -A_0 & (T-1)A_0 & -(T-1)A_0 \\ A_0 & (R-1)A_0 & -A_0 & -(R-1)A_0 \\ A_0 & -A_0 & -A_0 & A_0 \end{pmatrix} =$$

$$\begin{pmatrix} 1 & R-1 & T-1 & (T-1)(R-1) \\ 1 & -1 & T-1 & -(T-1) \\ 1 & R-1 & -1 & -(R-1) \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

IV. Main Theorem and Subsequent Theorems.

<u>Main Theorem</u>: $A_n C_n = M_n$.

<u>Lemma</u>: $A_n^2 = A_n A_n = \left| \prod_{i=1}^{n} N_i \right| I_{2^n}$ where I_{2^n} is the identity matrix of rank 2^n .

<u>Theorem</u>: $A_n^{-1} = \dfrac{A_n}{\prod_{i=1}^{n} N_i}$.

<u>Theorem</u>: $C_n = \dfrac{A_n M_n}{\prod_{i=1}^{n} N_i}$.

V. Computation Algorithm

<u>Definition</u>:

Let s_{ij} = 1 if $2i = j$ or $2i = j + 1$ when $1 \leq i \leq 2^{(n-1)}$;

= 0 if $2i < j$ or $2i > j + 1$ when $1 \leq i \leq 2^{(n-1)}$;

= $N_k - 1$ if $2^{(n-1)} + 2i = j$ when $2^{(n-1)} + 1 \leq i \leq 2^n$;

= -1 if $2^{(n-1)} + 2i = j + 1$ when $2^{(n-1)} + 1 \leq i \leq 2^n$;

= 0 if $2^{(n-1)} + 2i < j$ or $2^{(n-1)} + 2i > j + 1$ when $2^{(n-1)} + 1 \leq i \leq 2^n$.

Definition: Let $s_n^{(k)} = (s_{ij})$.

Theorem: $\prod_{k=1}^{n} S_n^{(k)} = A_n'$ (i.e. the transpose of A_n) .

This algorithm is similar to Yates'a.[2] It provides a simple way of calculating the transpose of A_n and of arriving at either the means squares or the covariances when the other is given.

We may derive all of the 64 equations in the four dimensional case readily by using the matrix for four dimensions:

$$A_3 = \begin{vmatrix} A_2 & (N_3-1)A_2 \\ A_2 & -A_2 \end{vmatrix} = \begin{vmatrix} \begin{vmatrix} A_1 & (N_2-1)A_1 \\ A_1 & -A_1 \end{vmatrix} & (N_3-1) & \begin{vmatrix} A_1 & (N_2-1)A_1 \\ A_1 & -A_1 \end{vmatrix} \\ \begin{vmatrix} A_1 & (N_2-1)A_1 \\ A_1 & -A_1 \end{vmatrix} & - & \begin{vmatrix} A_1 & (N_2-1)A_1 \\ A_1 & -A_1 \end{vmatrix} \end{vmatrix}$$

$$= \begin{vmatrix} A_1 & (N_2-1)A_1 & (N_3-1)A_1 & (N_3-1)(N_2-1)A_1 \\ A_1 & -A_1 & (N_3-1)A_1 & -(N_3-1)A_1 \\ A_1 & (N_2-1)A_1 & -A_1 & -(N_2-1)A_1 \\ A_1 & -A_1 & -A_1 & A_1 \end{vmatrix}$$

---

[2] F. Yates, The design and analysis of factorial experiments. Harpenden: Imperial Bureau of Soil Science, 1937. Pp. 15 and 29.

$$= \begin{vmatrix} \begin{vmatrix} A_0 & (N_1-1)A_0 \\ A_0 & -A_0 \end{vmatrix} & (N_2-1)\begin{vmatrix} A_0 & (N_1-1)A_0 \\ A_0 & -A_0 \end{vmatrix} & (N_3-1)\begin{vmatrix} A_0 & (N_1-1)A_0 \\ A_0 & -A_0 \end{vmatrix} & (N_3-1)(N_2-1)\begin{vmatrix} A_0 & (N_1-1)A_0 \\ A_0 & -A_0 \end{vmatrix} \\ \begin{vmatrix} A_0 & (N_1-1)A_0 \\ A_0 & -A_0 \end{vmatrix} & -\begin{vmatrix} A_0 & (N_1-1)A_0 \\ A_0 & -A_0 \end{vmatrix} & (N_3-1)\begin{vmatrix} A_0 & (N_1-1)A_0 \\ A_0 & -A_0 \end{vmatrix} & -(N_3-1)\begin{vmatrix} A_0 & (N_1-1)A_0 \\ A_0 & -A_0 \end{vmatrix} \\ \begin{vmatrix} A_0 & (N_1-1)A_0 \\ A_0 & -A_0 \end{vmatrix} & (N_2-1)\begin{vmatrix} A_0 & (N_1-1)A_0 \\ A_0 & -A_0 \end{vmatrix} & -\begin{vmatrix} A_0 & (N_1-1)A_0 \\ A_0 & -A_0 \end{vmatrix} & -(N_2-1)\begin{vmatrix} A_0 & (N_1-1)A_0 \\ A_0 & -A_0 \end{vmatrix} \\ \begin{vmatrix} A_0 & (N_1-1)A_0 \\ A_0 & -A_0 \end{vmatrix} & -\begin{vmatrix} A_0 & (N_1-1)A_0 \\ A_0 & -A_0 \end{vmatrix} & -\begin{vmatrix} A_0 & (N_1-1)A_0 \\ A_0 & -A_0 \end{vmatrix} & \begin{vmatrix} A_0 & (N_1-1)A_0 \\ A_0 & -A_0 \end{vmatrix} \end{vmatrix}$$

Since $A_0 = 1$, this is equal to the following matrix:

$$\begin{bmatrix} 1 & (N_1-1) & (N_2-1) & (N_2-1)(N_1-1) & (N_3-1) & (N_3-1)(N_1-1) & (N_3-1)(N_2-1) & (N_3-1)(N_2-1)(N_1-1) \\ 1 & -1 & (N_2-1) & -(N_2-1) & (N_3-1) & -(N_3-1) & (N_3-1)(N_2-1) & -(N_3-1)(N_2-1) \\ 1 & (N_1-1) & -1 & -(N_1-1) & (N_3-1) & (N_3-1)(N_1-1) & -(N_3-1) & -(N_3-1)(N_1-1) \\ 1 & -1 & -1 & 1 & (N_3-1) & -(N_3-1) & -(N_3-1) & (N_3-1) \\ 1 & (N_1-1) & (N_2-1) & (N_2-1)(N_1-1) & -1 & -(N_1-1) & -(N_2-1) & -(N_2-1)(N_1-1) \\ 1 & -1 & (N_2-1) & -(N_2-1) & -1 & 1 & -(N_2-1) & (N_2-1) \\ 1 & (N_1-1) & -1 & -(N_2-1) & -1 & -(N_1-1) & 1 & (N_1-1) \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix}$$

This matrix, in the case where the initial factor is individuals (I), becomes the following:

1	(R -1)	(T -1)	(T -1)(R -1)	(S -1)	(S -1)(R -1)	(S -1)(T -1)	(S -1)(T -1)(R -1)
1	-1	(T -1)	-(T -1)	(S -1)	-(S -1)	(S -1)(T -1)	-(S -1)(T -1)
1	(R -1)	-1	-(R -1)	(S -1)	(S -1)(R -1)	-(S -1)	-(S -1)(R -1)
1	-1	-1	1	(S -1)	-(S -1)	-(S -1)	(S -1)
1	(R -1)	(T -1)	(T -1)(R -1)	-1	-(R -1)	-(T -1)	-(T -1)(R -1)
1	-1	(T -1)	-(T -1)	-1	1	-(T -1)	(T -1)
1	(R -1)	-1	-(R -1)	-1	-(R -1)	1	(R -1)
1	-1	-1	1	-1	1	1	-1

In view of the matrix equation M = AC, we then have the following set of eight equations:

MS_i = A + (R-1)B + (T-1)C + (T-1)(R-1)D + (S-1)E + (S-1)(R-1)F + (S-1)(T-1)G + (S-1)(T-1)(R-1)H

$MS_{(ixr)}$ = A - B + (T-1)C - (T-1) D + (S-1)E - (S-1) F + (S-1)(T-1)G - (S-1)(T-1) H

$MS_{(ixt)}$ = A + (R-1)B - C - (R-1)D + (S-1)E + (S-1)(R-1)F - (S-1) G - (S-1) (R-1)H

$MS_{(ixrxt)}$ = A - B - C + D + (S-1)E - (S-1) F - (S-1) G + (S-1) H

$MS_{(ixs)}$ = A + (R-1)B + (T-1)C + (T-1)(R-1)D - E - (R-1)E - (T-1)G - (T-1)(R-1)H

$MS_{(ixrxs)}$ = A - B + (T-1)C - (T-1) D - E + F - (T-1)G + (T-1) H

$MS_{(ixtxs)}$ = A + (R-1)B - C - (R-1)D - E - (R-1)F + G + (R-1)H

$MS_{(ixrxtxs)}$ = A - B - C + D - E + F + G - H.

Now using the fact that the inverse of the matrix is the matrix itself divided by RST, we obtain the following solutions for the average variances and covariances:

$$\overline{s^2_{rts}} = \frac{1}{RTS}[P + (R-1)Q + (T-1)U + (T-1)(R-1)V + (S-1)W + (S-1)(R-1)X + (S-1)(T-1)Y + (S-1)(T-1)(R-1)Z]$$

$$\overline{cov(X_{rts},X_{r't's})} = \frac{1}{RTS}[P - Q + (T-1)U - (T-1)V + (S-1)W - (S-1)X + (S-1)(T-1)Y - (S-1)(T-1)Z]$$

$$\overline{cov(X_{rts},X_{rt's})} = \frac{1}{RTS}[P + (R-1)Q - U - (R-1)V + (S-1)W + (S-1)(R-1)X - (S-1)Y - (S-1)(R-1)Z]$$

$$\overline{cov(X_{rts},X_{r't's})} = \frac{1}{RTS}[P - Q - U + V + (S-1)W - (S-1)X - (S-1)Y + (S-1)Z]$$

$$\overline{cov(X_{rts},X_{rts'})} = \frac{1}{RTS}[P + (R-1)Q + (T-1)U + (T-1)(R-1)V - W - (R-1)X - (T-1)Y - (T-1)(R-1)Z]$$

$$\overline{cov(X_{rts},X_{r'ts'})} = \frac{1}{RTS}[P - Q + (T-1)U - (T-1)(R-1)V - W + X - (T-1)Y + (T-1)Z]$$

$$\overline{cov(X_{rts},X_{rt's'})} = \frac{1}{RTS}[P + (R-1)Q - U - (R-1)V - W - (R-1)X + Y + (R-1)Z]$$

$$\overline{cov(X_{rts},X_{r't's'})} = \frac{1}{RTS}[P - Q - U + V - W + X + Y - Z]$$

where $A = \overline{s^2_{rts}}$, $B = \overline{cov(X_{rts},X_{r'ts})}$ $C = \overline{cov(X_{rts},X_{rt's})}$ $D = \overline{cov(X_{rts},X_{r't's'})}$

$E = \overline{cov(X_{rts},X_{rts'})}$ $F = \overline{cov(X_{rts},X_{r'ts'})}$ $G = \overline{cov(X_{rts},X_{rt's'})}$, and $H = \overline{cov(X_{rts},X_{r't's'})}$;

$P = MS_i$ $Q = MS_{(ixr)}$ $U = MS_{(ixt)}$ $V = MS_{(ixrxt)}$ $W = MS_{(ixs)}$ $X = MS_{(ixrxs)}$ $Y = MS_{(ixtxs)}$

$Z = MS_{(ixrxtxs)}$

We may now specify the other 48 equations by interchanging letters in the matrix. This is so because the basic form of the matrix is the same when any factor is designated as the basic one. So to obtain the further equations we (for Table 1) interchange I and R,(for Table 2) I and T, and (for Table 3) I and S.

For Table 1 let:

$a = \overline{s^2_{its}}$ $\quad b = \overline{\mathrm{cov}(X_{its}, X_{i'ts})}$ $\quad c = \overline{\mathrm{cov}(X_{its}, X_{it's})}$ $\quad d = \overline{\mathrm{cov}(X_{its}, X_{i't's})}$

$e = \overline{\mathrm{cov}(X_{its}, X_{its'})}$ $\quad f = \overline{\mathrm{cov}(X_{its}, X_{i'ts'})}$ $\quad g = \overline{\mathrm{cov}(X_{its}, X_{it's'})}$

$h = \overline{\mathrm{cov}(X_{its}, X_{i't's'})}$ $\quad J = MS_r$ and the rest remain the same.

For Table 2 let

$\alpha = \overline{s^2_{ris}}$ $\quad \beta = \overline{\mathrm{cov}(X_{ris}, X_{r'is})}$ $\quad \gamma = \overline{\mathrm{cov}(X_{ris}, X_{ri's})}$

$\delta = \overline{\mathrm{cov}(X_{ris}, X_{r'i's})}$ $\quad \varepsilon = \overline{\mathrm{cov}(X_{ris}, X_{ris'})}$ $\quad \zeta = \overline{\mathrm{cov}(X_{ris}, X_{r'is'})}$

$\eta = \overline{\mathrm{cov}(X_{ris}, X_{ri's'})}$ $\quad \theta = \overline{\mathrm{cov}(X_{ris}, X_{r'i's'})}$ $\quad K = MS_t$ and the rest remain the same.

For Table 3 let

$\iota = \overline{s^2_{rti}}$ $\quad \kappa = \overline{\mathrm{cov}(X_{rti}, X_{r'ti})}$ $\quad \lambda = \overline{\mathrm{cov}(X_{rti}, X_{rt'i})}$ $\quad \mu = \overline{\mathrm{cov}(X_{rti}, X_{r't'i})}$

$\nu = \overline{\mathrm{cov}(X_{rti}, X_{rti'})}$ $\quad \xi = \overline{\mathrm{cov}(X_{rti}, X_{r'ti'})}$ $\quad o = \overline{\mathrm{cov}(X_{rti}, X_{rt'i'})}$

$\pi = \overline{\mathrm{cov}(X_{rti}, X_{r't'i'})}$ $\quad L = MS_s$ and the rest remain the same.

In addition to above: 1) $MS_{(rxt)} = M$ $\quad MS_{(rxs)} = N$ $\quad MS_{(rxtxs)} = 0$

2) $MS_{(txs)} = \rho$

We then obtain the following equations, shown in Tables 1-3:

Table 1

$$J = a + (I-1)b + (T-1)c + (T-1)(I-1)d + (S-1)e + (S-1)(I-1)f + (S-1)(T-1)g + (S-1)(T-1)(I-1)h$$

$$Q = a - b + (T-1)c - (T-1)d + (S-1)e - (S-1)f + (S-1)(T-1)g - (S-1)(T-1)h$$

$$M = a + (I-1)b - c - (I-1)d + (S-1)e + (S-1)(I-1)f - (S-1)g - (S-1)(I-1)h$$

$$V = a - b - c + d + (S-1)e - (S-1)f - (S-1)g + (S-1)h$$

$$N = a + (I-1)b + (T-1)c + (T-1)(I-1)d - e - (I-1)f - (T-1)g - (T-1)(I-1)h$$

$$X = a - b + (T-1)c - (T-1)d - e + f - (T-1)g + (T-1)h$$

$$O = a + (I-1)b - c - (I-1)d - e - (I-1)f + g + (I-1)h$$

$$Z = a - b - c + d - e + f + g - h$$

Inverse solution:

$$a = \frac{1}{ITS}[J + (I-1)Q + (T-1)M + (T-1)(I-1)V + (S-1)N + (S-1)(I-1)X + (S-1)(T-1)O + (S-1)(T-1)(I-1)Z]$$

$$b = \frac{1}{ITS}[J - Q + (T-1)M - (T-1)V + (S-1)N - (S-1)X + (S-1)(T-1)O - (S-1)(T-1)Z]$$

$$c = \frac{1}{ITS}[J + (I-1)Q - M - (I-1)V + (S-1)N + (S-1)(I-1)X - (S-1)O - (S-1)(I-1)Z]$$

$$d = \frac{1}{ITS}[J - Q - M + V + (S-1)N - (S-1)X - (S-1)O + (S-1)Z]$$

$$e = \frac{1}{ITS}[J + (I-1)Q + (T-1)M + (T-1)(I-1)V - N - (I-1)X - (T-1)O - (T-1)(I-1)Z]$$

$$f = \frac{1}{ITS}[J - Q + (T-1)M - (T-1)V - N + X - (T-1)O + (T-1)Z]$$

$$g = \frac{1}{ITS}[J + (I-1)Q - M - (I-1)V - N - (I-1)X + O + (I-1)Z]$$

$$h = \frac{1}{ITS}[J - Q - M + V - N + X + O - Z]$$

Table 2

$$K = \alpha + (R-1)\beta + (I-1)\gamma + (I-1)(R-1)\delta + (S-1)\varepsilon + (S-1)(R-1)\eta + (S-1)(T-1)\theta + (S-1)(I-1)(R-1)\lambda$$

$$M = \alpha - \beta + (I-1)\gamma - (I-1)\delta + (S-1)\varepsilon - (S-1)\eta + (S-1)(I-1)\theta - (S-1)(I-1)\lambda$$

$$U = \alpha + (R-1)\beta - \gamma - (R-1)\delta + (S-1)\varepsilon + (S-1)(R-1)\eta - (S-1)\theta - (S-1)(R-1)\lambda$$

$$V = \alpha - \beta - \gamma + \delta + (S-1)\varepsilon - (S-1)\eta - (S-1)\theta + (S-1)\lambda$$

$$\Gamma = \alpha + (R-1)\beta + (I-1)\gamma + (I-1)(R-1)\delta - \varepsilon - (R-1)\eta - (I-1)\theta - (I-1)(R-1)\lambda$$

$$O = \alpha - \beta + (I-1)\gamma - (I-1)\delta - \varepsilon + \eta - (I-1)\theta + (I-1)\lambda$$

$$Y = \alpha + (R-1)\beta - \gamma - (R-1)\delta - \varepsilon - (R-1)\eta + \theta + (R-1)\lambda$$

$$Z = \alpha - \beta - \gamma + \delta - \varepsilon + \eta + \theta - \lambda$$

Inverse solution:

$$\alpha = \frac{1}{RIS}[K + (R-1)M + (I-1)U + (I-1)(R-1)V + (S-1)\Gamma + (S-1)(R-1)O + (S-1)(I-1)Y + (S-1)(I-1)(R-1)Z]$$

$$\beta = \frac{1}{RIS}[K - M + (I-1)U - (I-1)V + (S-1)\Gamma - (S-1)O + (S-1)(I-1)Y - (S-1)(I-1)Z]$$

$$\gamma = \frac{1}{RIS}[K + (R-1)M - U - (R-1)V + (S-1)\Gamma + (S-1)(R-1)\Theta - (S-1)Y - (S-1)(R-1)Z]$$

$$\delta = \frac{1}{RIS}[K - M - U + V + (S-1)\Gamma - (S-1)O - (S-1)Y + (S-1)Z]$$

$$\varepsilon = \frac{1}{RIS}[K + (R-1)M + (I-1)U + (I-1)(R-1)V - \Gamma - (R-1)O - (I-1)Y - (I-1)(R-1)Z]$$

$$\eta = \frac{1}{RIS}[K - M + (I-1)U - (I-1)V - \Gamma + O - (I-1)Y + (I-1)Z]$$

$$\theta = \frac{1}{RIS}[K + (R-1)M - U - (R-1)V - \Gamma - (R-1)O + Y + (R-1)Z]$$

$$\lambda = \frac{1}{RIS}[K - M - U + V - \Gamma + O + Y - Z]$$

Table 3

$$L = \mu + (R-1)\nu + (T-1)\xi_1 + (T-1)(R-1)\pi + (I-1)\rho + (I-1)(R-1)\tau + (I-1)(T-1)\phi + (I-1)(T-1)(R-1)\psi$$

$$M = \mu - \nu + (T-1)\xi_1 - (T-1)\pi + (I-1)\rho - (I-1)\tau + (I-1)(T-1)\phi - (I-1)(T-1)\psi$$

$$T = \mu + (R-1)\nu - \xi_1 - (R-1)\pi + (I-1)\rho + (I-1)(R-1)\tau - (I-1)\phi - (I-1)(R-1)\psi$$

$$O = \mu - \nu - \xi_1 + \pi + (I-1)\rho - (I-1)\tau - (I-1)\phi + (I-1)\psi$$

$$W = \mu + (R-1)\nu + (T-1)\xi_1 + (T-1)(R-1)\pi - \rho - (R-1)\tau - (T-1)\phi - (T-1)(R-1)\psi$$

$$X = \mu - \nu + (T-1)\xi_1 - (T-1)\pi - \rho + \tau - (T-1)\phi + (T-1)\psi$$

$$Y = \mu + (R-1)\nu - \xi_1 - (R-1)\pi - \rho - (R-1)\tau + \phi + (R-1)\psi$$

$$Z = \mu - \nu - \xi_1 + \pi - \rho + \tau - \phi + \psi$$

Inverse solution:

$$\mu = \frac{1}{RTI}[L + (R-1)M + (T-1)N + (T-1)(R-1)O + (I-1)W + (I-1)(R-1)X + (I-1)(T-1)Y + (I-1)(T-1)(R-1)Z]$$

$$\nu = \frac{1}{RTI}[L - M + (T-1)N - (T-1)O + (I-1)W - (I-1)X + (I-1)(T-1)Y - (I-1)(T-1)Z]$$

$$\xi_1 = \frac{1}{RTI}[L + (R-1)M - N - (R-1)O + (I-1)W + (I-1)(R-1)X - (I-1)Y - (I-1)(R-1)Z]$$

$$\pi = \frac{1}{RTI}[L - M - N + O + (I-1)W - (I-1)X - (I-1)Y + (I-1)Z]$$

$$\rho = \frac{1}{RTI}[L + (R-1)M + (T-1)N + (T-1)(R-1)O - W - (R-1)X - (T-1)Y - (T-1)(R-1)Z]$$

$$\tau = \frac{1}{RTI}[L - M + (T-1)N - (T-1)O - W + X - (T-1)Y + (T-1)Z]$$

$$\phi = \frac{1}{RTI}[L + (R-1)M - N - (R-1)O - W + X + Y - (R-1)Z]$$

$$\psi = \frac{1}{RTI}[L - M - N + O + W - X - Y + Z]$$

We may derive the five dimensional matrix in the same way as the four dimensional one. This matrix appears in Table 4. From this matrix we may find the 16 equations relating 16 of the mean squares (those which relate to i) with 16 of the average variances and covariances (those which are based on i). Let the following letters represent the designated mean squares and average variance and covariances:

$A = MS_i$

$B = MS_{(i \times r)}$

$C = MS_{(i \times t)}$

$D = MS_{(i \times r \times t)}$

$E = MS_{(i \times s)}$

$F = MS_{(i \times r \times s)}$

$G = MS_{(i \times t \times s)}$

$H = MS_{(i \times r \times t \times s)}$

$J = MS_{(i \times m)}$

$K = MS_{(i \times r \times m)}$

$M = MS_{(i \times t \times m)}$

$O = MS_{(i \times r \times t \times m)}$

$P = MS_{(i \times s \times m)}$

$Q = MS_{(i \times r \times s \times m)}$

$U = MS_{(i \times t \times s \times m)}$

$V = MS_{(i \times r \times t \times s \times m)}$

$a = \overline{s^2_{rtsm}}$

$b = \overline{cov(X_{rtsm}, X_{r'tsm})}$

$c = \overline{cov(X_{rtsm}, X_{rt'sm})}$

$d = \overline{cov(X_{rtsm}, X_{r't'sm})}$

$e = \overline{cov(X_{rtsm}, X_{rts'm})}$

$f = \overline{cov(X_{rtsm}, X_{r'ts'm})}$

$k = \overline{cov(X_{rtsm}, X_{rt's'm})}$

$h = \overline{cov(X_{rtsm}, X_{r't's'm})}$

$j = \overline{cov(X_{rtsm}, X_{rtsm'})}$

$k = \overline{cov(X_{rtsm}, X_{r'tsm'})}$

$n = \overline{cov(X_{rtsm}, X_{rt'sm'})}$

$o = \overline{cov(X_{rtsm}, X_{r't'sm'})}$

$p = \overline{cov(X_{rtsm}, X_{rts'm'})}$

$q = \overline{cov(X_{rtsm}, X_{r'ts'm'})}$

$u = \overline{cov(X_{rtsm}, X_{rt's'm'})}$

$v = \overline{cov(X_{rtsm}, X_{r't's'm'})}$

The 32 equations relating the mean squares and the average variance and covariances follow in Tables 5 and 6.

Table 4

1	(R-1)	(T-1)	(T-1)(R-1)	(S-)	(S-1)(R-1)	(S-1)(T-1)	(S-1)(T-1)(R-1)	(M-1)	(M-1)(R-1)	(M-1)(T-1)		
1	-1	(T-1)	-(T-1)	(S-)	-(S-1)	(S-1)(T-1)	-(S-1)(T-1)	(M-1)	-(M-1)	(M-1)(T-1)		
1	(R-1)	-1	-(R-1)	(S-)	(S-1)(R-1)	-(S-1)	-(S-1)(R-1)	(M-1)	(M-1)(R-1)	-(M-1)		
1	-1	-1	1	(S-)	-(S-1)	-(S-1)	(S-1)	(M-1)	-(M-1)	-(M-1)		
1	(R-1)	(T-1)	(T-1)(R-1)	-1	-(R-1)	-(T-1)	-(T-1)(R-1)	(M-1)	(M-1)(R-1)	(M-1)(T-1)		
1	-1	(T-1)	-(T-1)	-1	1	-(T-1)	(T-1)	(M-1)	-(M-1)	(M-1)(T-1)		
1	(R-1)	-1	-(R-1)	-1	-(R-1)	1	(R-1)	(M-1)	(M-1)(R-1)	-(M-1)		
1	-1	-1	1	-1	1	1	-1	(M-1)	-(M-1)	-(M-1)		
1	(R-1)	(T-1)	(T-1)(R-1)	(S-1	(S-1)(R-1)	(S-1)(T-1)	(S-1)(T-1)(R-1)	-1	-(R-1)	-(T-1)		
1	-1	(T-1)	-(T-1)	(S-1	-(S-1)	(S-1)(T-1)	-(S-1)(T-1)	-1	1	-(T-1)		
1	(R-1)	-1	-(R-1)	(S-1	(S-1)(R-1)	-(S-1)	-(S-1)(R-1)	-1	-(R-1)	1		
1	-1	-1	1	(S-1	-(S-1)	-(S-1)	(S-1)	-1	1	1		
1	(R-1)	(T-1)	(T-1)(R-1)	-1	-(R-1)	-(T-1)	-(T-1)(R-1)	-1	-(R-1)	-(T-1)		
1	-1	(T-1)	-(T-1)	-1	1	-(T-1)	(T-1)	-1	1	-(T-1)		
1	(R-1)	-1	-(R-1)	-1	-(R-1)	1	(R-1)	-1	-(R-1)	1		
1	-1	-1	1	-1	1	1	-1	-1	1	1		

Table 4 Cont.

(M-1)(T-1)(R-1)	(M-1)(S-1)	(M-1)(S-1)(R-1)	(M-1)(S-1)(T-1)	(M-1)(S-1)(T-1)(R-1)
-(M-1)(T-1)	(M-1)(S-1)	-(M-1)(S-1)	(M-1)(S-1)(T-1)	-(M-1)(S-1)(T-1)
-(M-1)(R-1)	(M-1)(S-1)	(M-1)(S-1)(R-1)	-(M-1)(S-1)	-(M-1)(S-1)(R-1)
(M-1)	(M-1)(S-1)	-(M-1)(S-1)	-(M-1)(S-1)	(M-1)(S-1)
(M-1)(T-1)(R-1)	-(M-1)	-(M-1)(R-1)	-(M-1)(T-1)	-(M-1)(T-1)(R-1)
-(M-1)(T-1)	-(M-1)	(M-1)	-(M-1)(T-1)	(M-1)(T-1)
-(M-1)(R-1)	-(M-1)	-(M-1)(R-1)	(M-1)	(M-1)(R-1)
(M-1)	-(M-1)	(M-1)	(M-1)	-(M-1)
-(T-1)(R-1)	-(S-1)	-(S-1)(R-1)	-(S-1)(T-1)	-(S-1)(T-1)(R-1)
(T-1)	-(S-1)	(S-1)	-(S-1)(T-1)	(S-1)(T-1)
(R-1)	-(S-1)	-(S-1)(R-1)	(S-1)	(S-1)(R-1)
-1	-(S-1)	(S-1)	(S-1)	-(S-1)
-(T-1)(R-1)	1	(R-1)	(T-1)	(T-1)(R-1)
(T-1)	1	-1	(T-1)	-(T-1)
(R-1)	1	(R-1)	-1	-(R-1)
-1	1	-1	-1	1

This matrix is a continuation of the one in Table 4 .

Table 5

A = a + (R-1)b + (T-1)c + (T-1)(R-1)d + (S-1)e + (S-1)(R-1)f + (S-1)(T-1)g + (S-1)(T-1)(R-1)h + (M-1)j + (M-1)(R-1)k
+ (M-1)(T-1)m + (M-1)(T-1)(R-1)o + (M-1)(S-1)p + (M-1)(S-1)(R-1)q + (M-1)(S-1)(T-1)u + (M-1)(S-1)(T-1)(R-1)v,

B = a - b + (T-1)c - (T-1)d + (S-1)e - (S-1)f + (S-1)(T-1)g - (S-1)(T-1)h + (M-1)j - (M-1)k
+ (M-1)(T-1)m - (M-1)(T-1)o + (M-1)(S-1)p - (M-1)(S-1)q + (M-1)(S-1)(T-1)u - (M-1)(S-1)(T-1)v,

C = a + (R-1)b - c - (R-1)d + (S-1)e + (S-1)(R-1)f - (S-1)g - (S-1)(R-1)h + (M-1)j + (M-1)(R-1)k
- (M-1)n - (M-1)(R-1)o + (M-1)(S-1)p + (M-1)(S-1)(R-1)q - (M-1)(S-1)u - (M-1)(S-1)(R-1)v,

D = a - b - c + d + (S-1)e - (S-1)f - (S-1)g + (S-1)h + (M-1)j - (M-1)k
- (M-1)n + (M-1)o + (M-1)(S-1)p - (M-1)(S-1)q - (M-1)(S-1)u + (M-1)(S-1)v

E = a + (R-1)b + (T-1)c + (T-1)(R-1)d - e - (R-1)f - (T-1)g - (T-1)(R-1)h + (M-1)j + (M-1)(R-1)k
+ (M-1)(T-1)n + (M-1)(T-1)(R-1)o - (M-1)p - (M-1)(R-1)q - (M-1)(T-1)u - (M-1)(T-1)(R-1)v,

F = a - b + (T-1)c - (T-1)d - e + f - (T-1)g + (T-1)h + (M-1)j - (M-1)k
+ (M-1)(T-1)n - (M-1)(T-1)o - (M-1)p - (M-1)q - (M-1)(T-1)u + (M-1)(T-1)v,

G = a + (R-1)b - c - (R-1)d - e - (R-1)f + g + (R-1)h + (M-1)j + (M-1)(R-1)k
- (M-1)n - (M-1)(R-1)o - (M-1)p - (M-1)(R-1)q + (M-1)u + (M-1)(R-1)v.

H = a - b - c + d - e + f + g - h + (M-1)j - (M-1)k
- (M-1)n + (M-1)o + (M-1)p + (M-1)q + (M-1)u - (M-1)v.

Table 5 Cont.

$$J = a + (R-1)b + (T-1)c + (T-1)(R-1)d + (S-1)e + (S-1)(R-1)f + (S-1)(T-1)g + (S-1)(T-1)(R-1)h$$
$$-j - (R-1)k - (T-1)n - (T-1)(R-1)o - (S-1)p - (S-1)(R-1)q - (S-1)(T-1)u - (S-1)(T-1)(R-1)v,$$

$$K = a - b + (T-1)c - (T-1)d + (S-1)e - (S-1)f + (S-1)(T-1)g - (S-1)(T-1)h$$
$$-j + k - (T-1)n + (T-1)o - (S-1)p + (S-1)q - (S-1)(T-1)u + (S-1)(T-1)v,$$

$$M = a + (R-1)b - c - (R-1)d + (S-1)e + (S-1)(R-1)f - (S-1)g - (S-1)(R-1)h$$
$$-j - (R-1)k + n + (R-1)o - (S-1)p - (S-1)(R-1)q + (S-1)u + (S-1)(R-1)v,$$

$$O = a - b - c + d + (S-1)e - (S-1)f - (S-1)g + (S-1)h$$
$$-j + k + n - o - (S-1)p + (S-1)q + (S-1)u - (S-1)v,$$

$$P = a + (R-1)b + (T-1)c + (T-1)(R-1)d - e - (R-1)f - (T-1)g - (T-1)(R-1)h$$
$$-j - (R-1)k - (T-1)n - (T-1)(R-1)o + p + (R-1)q + (T-1)u + (T-1)(R-1)v,$$

$$Q = a - b + (T-1)c - (T-1)d - e + f - (T-1)g + (T-1)h$$
$$-j + k - (T-1)u + (T-1)o + p - q + (T-1)u - (T-1)v,$$

$$U = a + (R-1)b - c - (R-1)d - e - (R-1)f + g + (R-1)h$$
$$-j - (R-1)k + n + (R-1)o + p + (R-1)q - u - (R-1)v,$$

$$V = a - b - c + d - e + f + g - h$$
$$-j + k + n - o + p - q - u + v.$$

Table 6

$$a = \frac{1}{RSTM}\Big[A + (R-1)B + (T-1)C + (T-1)(R-1)D + (S-1)E + (S-1)(R-1)F + (S-1)(T-1)G + (S-1)(T-1)(R-1)H + (M-1)J +$$
$$(M-1)(R-1)K + (M-1)(T-1)N + (M-1)(T-1)(R-1)O + (M-1)(S-1)P + (M-1)(S-1)(R-1)Q + (M-1)(S-1)(T-1)U + (M-1)(S-1)(T-1)(R-1)\overline{V}\Big]$$

$$b = \frac{1}{RSTM}\Big[A - B + (T-1)C - (T-1)D + (S-1)E - (S-1)F + (S-1)(T-1)G - (S-1)(T-1)H + (M-1)J -$$
$$(M-1)K + (M-1)(T-1)N - (M-1)(T-1)O + (M-1)(S-1)P - (M-1)(S-1)Q + (M-1)(S-1)(T-1)U - (M-1)(S-1)(T-1)\overline{V}\Big]$$

$$c = \frac{1}{RSTM}\Big[A + (R-1)B - C - (R-1)D + (S-1)E + (S-1)(R-1)F - (S-1)G - (S-1)(R-1)H + (M-1)J +$$
$$(M-1)(R-1)K - (M-1)N - (M-1)(R-1)O + (M-1)(S-1)P + (M-1)(S-1)(R-1)Q - (M-1)(S-1)U - (M-1)(S-1)(R-1)\overline{V}\Big]$$

$$d = \frac{1}{RSTM}\Big[A - B - C + D + (S-1)E - (S-1)F - (S-1)G + (S-1)H + (M-1)J -$$
$$(M-1)K - (M-1)N + (M-1)O + (M-1)(S-1)P - (M-1)(S-1)Q - (M-1)(S-1)U + (M-1)(S-1)\overline{V}\Big]$$

$$e = \frac{1}{RSTM}\Big[A + (R-1)B + (T-1)C + (T-1)(R-1)D - E - (R-1)F - (T-1)G - (T-1)(R-1)H + (M-1)J +$$
$$(M-1)(R-1)K + (M-1)(T-1)N - (M-1)(T-1)(R-1)O - (M-1)P - (M-1)(R-1)Q - (M-1)(T-1)U - (M-1)(T-1)(R-1)\overline{V}\Big]$$

$$f = \frac{1}{RSTM}\Big[A - B + (T-1)C - (T-1)D - E + F - (T-1)G + (T-1)H + (M-1)J -$$
$$(M-1)K + (M-1)(T-1)N - (M-1)(T-1)O - (M-1)P + (M-1)Q - (M-1)(T-1)U + (M-1)(T-1)\overline{V}\Big]$$

$$g = \frac{1}{RSTM}\Big[A + (R-1)B - C - (R-1)D - E - (R-1)F + G + (R-1)H + (M-1)J +$$
$$(M-1)(R-1)K - (M-1)N - (M-1)(R-1)O - (M-1)P - (M-1)(R-1)Q + (M-1)U + (M-1)(R-1)\overline{V}\Big]$$

$$h = \frac{1}{RSTM}\Big[A - B - C + D - E + F + G - H + (M-1)J -$$
$$(M-1)K - (M-1)N + (M-1)O - (M-1)P + (M-1)Q + (M-1)U - (M-1)\overline{V}\Big]$$

Table 6 Cont.

$$j = \frac{1}{RSTM}\Big[A + (R-1)B + (T-1)C + (T-1)(R-1)D + (S-1)E + (S-1)(R-1)F + (S-1)(T-1)G + (S-1)(T-1)(R-1)H$$
$$-J - (R-1)K - (T-1)N - (T-1)(R-1)O - (S-1)P - (S-1)(R-1)Q - (S-1)(T-1)U - (S-1)(T-1)(R-1)V\Big]$$

$$k = \frac{1}{RSTM}\Big[A - B + (T-1)C - (T-1)D + (S-1)E - (S-1)F + (S-1)(T-1)G - (S-1)(T-1)H$$
$$-J + K - (T-1)N + (T-1)O - (S-1)P + (S-1)Q - (S-1)(T-1)U + (S-1)(T-1)V\Big]$$

$$n = \frac{1}{RSTM}\Big[A + (R-1)B - C - (R-1)D + (S-1)E + (S-1)(R-1)F - (S-1)G - (S-1)(R-1)H$$
$$-J - (R-1)K + N + (R-1)O - (S-1)P - (S-1)(R-1)Q + (S-1)U + (S-1)(R-1)V\Big]$$

$$o = \frac{1}{RSTM}\Big[A - B - C + D + (S-1)E - (S-1)F - (S-1)G + (S-1)H$$
$$-J + K + N - O - (S-1)P + (S-1)Q + (S-1)U - (S-1)V\Big]$$

$$p = \frac{1}{RSTM}\Big[A + (R-1)B + (T-1)C + (T-1)(R-1)D - E - (R-1)F - (T-1)G - (T-1)(R-1)H$$
$$-J - (R-1)K - (T-1)N - (T-1)(R-1)O + P + (R-1)Q + (T-1)U + (T-1)(R-1)V\Big]$$

$$q = \frac{1}{RSTM}\Big[A - B + (T-1)C - (T-1)D - E + F - (T-1)G + (T-1)H$$
$$-J + K - (T-1)N + (T-1)O + P - Q + (T-1)U - (T-1)V\Big]$$

$$u = \frac{1}{RSTM}\Big[A + (R-1)B - C - (R-1)D - B - (R-1)F + G + (R-1)H$$
$$-J - (R-1)H + N + (R-1)O + P + (R-1)Q - U - (R-1)V\Big]$$

$$v = \frac{1}{RSTM}\Big[A - B - C + D - E + F + G - H$$
$$-J + K + N - O + P - Q + U + V\Big]$$

Chapter 4. Application of the Technique

For the purposes of illustrating this technique in the four factor case, some data has been taken from the Johnson-Stanley study.[1] In the original experiment an analysis of the data was carried out according to a four-factor factorial design with two replications. Upon reanalyzing we pooled two of the factors and artificialized the replication into another factor. The new design is a factorial design with four factors: 40 individuals (I), 2 sexes (S), 4 levels of H (the combined factor), and 2 levels of R (the factor obtained from replication).

When the newly organized data were analyzed in a standard factorial analysis of variance, the results in Table 7 were obtained. The appropriate F-ratios were determined by the expected-mean-square breakdown (for I and R random, H and S fixed) given in the Appendix.

In accordance with the theoretical results expressed above, we see that the relationships between mean squares and mean covariances may be represented as in the formulae derived in the previous section.

We get all of these equations through a rotation of the arrays, i.e. we let each factor in turn represent the initial or reference factor (rows). From this we obtain the values in Tables 8, 9, 10, and 11 for the mean covariances and their respective estimates of the mean correlations.

As an illustration of the calculation of one of these mean covariances, let us take the second expression on p. 31:

$$\overline{\text{cov}(X_{rsh}, X_{r'sh})} = [(MS)_i - (MS)_{ir} + (H-1)(MS)_{ih} - (H-1)(MS)_{irh} + (S-1)(MS)_{is} - (S-1)(MS)_{irs} + (H-1)(S-1)(MS)_{ish} - (H-1)(S-1)(MS)_{irsh}]/RSH$$

---

[1] O. G. Johnson and J. C. Stanley, Attitudes toward authority of delinquent and non-delinquent boys, J. abnorm. soc. Psychol., 1955, 51, 712-716.

Table 7

Analysis of Variance (I,R random; H,S fixed)

Source of Variation	df	Mean Square	Error Term	Source of Error Term	d.f.	Estimated Variance Component
i	39	10.1698****	1.7560	$MS_{(i \times r)}$	39	0.526
s	1	12.9391	16.6410	$MS_{(i \times s)} + MS_{(s \times r)} - MS_{(i \times s \times r)}$	1.01	
h	3	2.1182	5.4676	$MS_{(i \times h)} + MS_{(h \times r)} - MS_{(i \times h \times r)}$	4.25	
r	1	1.7016	1.7560	$MS_{(i \times r)}$ +	39	
i x s	39	0.9006	0.8362	$MS_{(i \times s \times r)}$	39	0.008
i x h	117	1.8255****	0.8470	$MS_{(i \times h \times r)}$	117	0.245
i x r	39	1.7560*	1.1431	$MS_{(i \times s \times h \times r)}$ +	117	≥0.079
s x h	3	1.1932	3.4037	$MS_{(i \times s \times h)} + MS_{(s \times h \times r)} - MS_{(i \times s \times h \times r)}$	3.22	
s x r	1	16.5766****	0.8362	$MS_{(i \times s \times r)}$	39	0.098
h x r	3	4.4891***	0.8470	$MS_{(i \times h \times r)}$	117	0.046
i x s x h	117	1.2744	1.1431	$MS_{(i \times s \times h \times r)}$ +	117	0.066
i x s x r	39	0.8362	1.1431	$MS_{(i \times s \times h \times r)}$ +	117	
i x h x r	117	0.8470	1.1431	$MS_{(i \times s \times h \times r)}$ +	117	
s x h x r	3	3.7724	1.1431	$MS_{(i \times s \times h \times r)}$ +	117	0.053
i x s x h x r	117	1.1431				
Total	639					

+ Assuming $\sigma^2_{ishr} = 0$

Significance levels: (* .05), (** .01),), (*** .005), (**** .001)

TABLE 8

Empirical values of average variances and covariances

(Johnson-Stanley Data)

	Empirical value	Corresponding estimated average correlation
$\overline{s^2_{rsh}}$	1.6083	
$\overline{cov(X_{rsh}, X_{r'sh})}$	0.7380	.408
$\overline{cov(X_{rsh}, X_{rs'h})}$	0.6846	.379
$\overline{cov(X_{rsh}, X_{rsh'})}$	0.5358	.296
$\overline{cov(X_{rsh}, X_{r's'h})}$	0.6807	.376
$\overline{cov(X_{\ \ .}, X_{\ \ .})}$	0.4605	.255
$\overline{cov(X_{rsh}, X_{r'sh'})}$	0.4605	.255
$\overline{cov(X_{rsh}, X_{rs'h'})}$	0.6209	.343
$\overline{cov(X_{rsh}, X_{r's'h'})}$	0.4689	.259

TABLE 9

Empirical values of average variances and covariances

	Empirical value	Corresponding estimated average correlation
$\overline{s^2_{ish}}$	1.1734	
$\overline{cov(X_{ish}, X_{i'sh})}$	0.1031	0.0880
$\overline{cov(X_{ish}, X_{is'h})}$	-0.0312	-0.0266
$\overline{cov(X_{ish}, X_{ish'})}$	0.1062	0.0905
$\overline{cov(X_{ish}, X_{i's'h})}$	-0.0258	-0.0220
$\overline{cov(X_{ish}, X_{i'sh'})}$	0.0310	0.0264
$\overline{cov(X_{ish}, X_{is'h'})}$	0.0970	0.0834
$\overline{cov(X_{ish}, X_{i's'h'})}$	-0.0541	-0.0461

TABLE 10

Empirical values of average variances and covariances

	Empirical value	Corresponding estimated average correlation
$\overline{s^2_{irs}}$	1.3099	
$\overline{cov(X_{irs}, X_{i'rs})}$	0.0374	0.0285
$\overline{cov(X_{irs}, X_{ir's})}$	0.2632	0.2009
$\overline{cov(X_{irs}, X_{irs'})}$	0.0755	0.0577
$\overline{cov(X_{irs}, X_{i'r's})}$	-0.0347	-0.0265
$\overline{cov(X_{irs}, X_{i'rs'})}$	0.0118	0.0090
$\overline{cov(X_{irs}, X_{ir's'})}$	0.2047	0.1563
$\overline{cov(X_{irs}, X_{i'r's'})}$	-0.0071	-0.0054

TABLE 11

Empirical values of average variances and covariances

	Empirical value	Corresponding estimated average correlation
$\overline{\sigma^2_{irh}}$	1.2297	
$\overline{cov(X_{irh}, X_{i'rh})}$	0.0807	0.0656
$\overline{cov(X_{irh}, X_{ir'h})}$	0.0250	0.0203
$\overline{cov(X_{irh}, X_{i'r'h})}$	-0.0323	-0.0263
$\overline{cov(X_{irh}, X_{irh'})}$	-0.0047	-0.0038
$\overline{cov(X_{irh}, X_{i'rh'})}$	0.0804	0.0654
$\overline{cov(X_{irh}, X_{ir'h'})}$	-0.0130	-0.0106
$\overline{cov(X_{irh}, X_{i'r'h'})}$	-0.0047	-0.0038

$$= [10.1698 - 1.7560 + (3)(1.8255) - (3)(0.8470) + (1)(0.9006) - (1)(0.8362) - (3)(1)(1.2744) - (3)(1)(1.431)]/(2)(2)(4)$$

$$= 11.8076/16 = 0.7380$$

By similar calculation we find that $\overline{s^2_{res}} = 1.8083$. We therefore estimate the average intercorrelation across replications, within sex, and within the levels of the 'H' factor to be 0.7380/1.8083 = 0.408. This is the highest of all the estimated r's, as indeed it should be if Replicate 1 is similar to Replicate 2. Note how well this .408 agrees with the directly computed $\bar{r}$ of .41 reported by Johnson and Stanley.[2] Also, the weighted mean of .379, .296, ...259 in Table 2a is .318, which agrees with the .32 estimated by Johnson and Stanley[3] in another way.

Another method of teacher rating is illustrated by Cosgrove's forced-choice teacher rating scale. This technique was applied by Newell[4] in a study of differing instructional approaches, where he used Cosgrove's rating form, which consists of ten sets of four descriptive phrases each. One phrase in each set is supposed to be representative of a particular factor found in a factor analysis of an original item pool. The students were supposed to rank the phrases in each set in an ipsative fashion for applicability to their instructor. Sixteen students rated each instructor.

This procedure outlined above precludes any standard analysis of variance, because the highest order interaction sum of squares is a

---

[2] Ibid., p. 714.

[3] Ibid.

[4] J. M. Newell, Experimental verification of three instructional approaches. Mimeographed, 1961.

linear function on the other sums of squares, and therefore the mean squares are not independent. This is true of any forced-choice ranking scheme. This non-independance of mean squares does not, however, invalidate the procedure for finding the average variance and covariances. This procedure is valid for any array of real numbers regardless of their form. Obviously, however, the rating procedure partially fixes the correlational structure of the data and this is reflected in the mean squares via the average variance and covariances.

In the Johnson-Stanley experiment there was no forced-choice rating. In every situation where a rating occurred the judges were free to make as many choices of a given type as they preferred. Thus, the Johnson-Stanley experiment was amenable to analysis of variance treatment, no further correlational relationships having been imposed on the data.

Chapter 5. Rating Teachers

If one desires to evaluate teachers in various places ($p = 1, 2, \ldots, P$) with respect to classroom performance, he can of course have them all rated on a number of traits by raters ($r = 1, 2, \ldots, R$) who visit each teacher's classroom either individually or as a team on Q occasions ($o = 1, 2, \ldots, O$). Teachers are usually "nested" within place, for they seldom teach in more than one school during a given academic year. The detailed sampling plan for this kind of rating study must be developed carefully in advance. Shall one merely choose a teacher at random from the total group of teachers and then choose randomly an occasion (day and time) for that teacher? Or shall one, instead, choose a place at random, then a teacher or teachers at random, and then occasions randomly? Probably some places have many teachers, others few. Within-place and between-occasion variability may be relevant considerations in setting up the sampling frame. We shall not elaborate on the technical aspects, for they are treated extensively in a number of places.[1]

If in a rating study there are P places, I_p instructors in each place, R raters, T traits, and S ratings of each instructor by each rater on each trait, all assigned wholly at random, then one may use the following linear model for the rating X_{pirts} of the ith instructor in the pth place by the rth rater on the tth trait the sth time:

$$X_{pirts} = \Theta + \alpha_p + A_{i(p)} + \gamma_r + \beta_t + \varepsilon_s + \zeta_{pr} + \eta_{pt} + \iota_{ps} + \kappa_{rt} + \lambda_{rs} + \mu_{ts} + \nu_{ir(p)} + \xi_{it(p)} + o_{is(p)} + \pi_{prt} + \rho_{prs} + \sigma_{pts} + \tau_{rts} + \upsilon_{irt(p)} + \phi_{irs(p)} + \chi_{its(p)} + \psi_{prts} + \omega_{irts(p)} + E_{pirts}].$$

[1] For example, see W. Deming, Sample Design in Business Research. New York: Wiley, 1960.

where $\ominus$ represents the "general contribution," estimated by the study's mean, $\bar{X}_{.....}$.

Perhaps p, i, and r may be considered random effects, for if we repeated the study we would use a fresh batch of places, instructors, and raters "like those," so we are "drawing" places, instructors, and raters conceptually in a random fashion from three parent populations. It seems likely, however, that in repeating the study we would use the same traits, since probably they were chosen on a judgmental, non-random basis. Thus traits would constitute a fixed effect. Also, we would again use the 1, 2, ..., s ratings, so the s factor is fixed. The model, then, is mixed, 3 random effects to 2 fixed effects. This complicates the analysis, but the faithful mirroring of "reality" may be worth the effort. Assumptions are as set forth by Cornfield and Tukey,[2] and modified for nested designs by Green and Tukey and by Stanley.[3]

The partitioning-of-covariances procedures featured in this report are applicable to the above design. Also, one may treat the instructors as a fixed effect and adjust their ratings to neutralize eleven of the twelve possible sources of rater bias statistically: r, p x r, r x t, r x s, i x r, p x r x t, p x r x s, r x t x s, i x r x t, i x r x s, and p x r x t x s. This is complex and tedious, but possible. How much the reliability of the ratings has been affected by the adjustments can be determined by comparing $\overline{\text{cov}(X_{\text{pirts}}, X_{\text{pirts}'})} \Big/ \overline{\text{cov}(X_{\text{pirts}}, X_{\text{pirts}})}$ with $\overline{\text{cov}(X'_{\text{pirts}}, X'_{\text{pirts}'})} \Big/ \overline{\text{cov}(X'_{\text{pirts}}, X'_{\text{pirts}})}$, where primes denote adjusted ratings.

---

[2] J. Cornfield and J. W. Tukey, Average values of mean squares in factorials, Ann. Math. Statist., 1956, 27, 907-949.

[3] B. F. Green, Jr., and J. W. Tukey, Complex analyses of variance: general problems. Psychometrika, 1960, 25, 127-152.

J. C. Stanley, Analysis of a doubly nested design, Educ. psychol. Measmt., 1961, 21, 831-839; 1962, 72, 11.

The adjustment formula is

$$X'_{pirta} = X_{pirta} - (1 - \sqrt{1/F_r})(\bar{X}_{..r..} - \bar{X}_{.....}) - (1 - \sqrt{1/F_{(p \times r)}})(\bar{X}_{.p.r..} - \bar{X}_{.p....} - \bar{X}_{..r..} + \bar{X}_{.....}) - \ldots,$$

where F_r is the ratio of MS_r to its appropriate "error" MS, etc. The adjustment changes MS_r to the magnitude of its error term, $MS_{(p \times r)}$ to the size of its error term, and so on. Such adjustments may be made legitimately even if the F involved is not significant, but they are needed only when the particular bias being adjusted for is fairly strong.

Incomplete designs. Because of limited time or money, it will not always be feasible for every rater to visit every classroom. Also, it will usually be desirable for a least two raters to be rating the same teacher independently at the same time, in order that inter-rater consistency within occasion may be evaluated. One might, for example, have R raters available and I instructors to be rated on T traits. Two raters could be assigned at random to rate each of two instructors on all T traits twice per occasion on each of two randomly drawn occasions, using 2T randomly ordered rating cards per rater on each occasion. Also, every instructor would be rated on four separate occasions by each of two separate teams of raters. The R available raters would constitute the R/2 rating teams. Thus, each instructor would be visited four times and rated four times during each visit on each trait, yielding a total of 16 sets of ratings per instructor, given by four raters. (Note the various ways in which replication occurs in this design.)

This leads to a random pair of instructors, say 1 and 2, each of whom is rated by Teams 1 and 2. Team 1 consists of randomly assigned Raters 1 and 2, Team 2 of Raters 3 and 4. Instructors and teams are nested within pair but cross each other. Raters are nested within team, but cross instructors within pair. The replication of rating within

occasion, the 2 occasions, and the T traits cross everything. The considerable amount of nesting reduces the number of higher-order interactions greatly, thereby simplifying the analysis. Instructors can interact with rating teams within pairs, and instructors can interact with raters whith rating teams within pairs, but the remaining 13 interactions involving pair, instructor, team, and/or rater cannot be estimated in this design.

The above outline is suggestive only. A huge number of possible designs exist, so that the objectives and resources of a given investigation can be the guiding basis for setting up experiments expertly if the researcher is aware of the many possibilities. One should of course tailor the design to fit the problem, not vice versa. Little use has yet been made in educational research of, for example, partially balanced incomplete block designs, even though they offer needed flexibility in many situations. Recently, Page[4] commented on this unfortunate lag in utilization of principles and techniques well known in other fields. The serious researcher studying teacher merit will want to familiarize himself with several expositions of such designs. The most comprehensive "cookbook" is probably still the Cochran and Cox standard.[5] Cox[6] offers a simple but careful treatment. Also see the books by Finney and by Freund.[7]

---

[4]E. B. Page, A flexible incomplete block design for classroom experiments. Amer. Psychologist, 1961, 16, 432. Abstract.

[5]W. G. Cochran and Gertrude M. Cox, Experimental designs (rev. ed.). New York: Wiley, 1957.

[6]D. R. Cox, Planning of experiments. New York: Wiley, 1958.

[7]D. J. Finney, The theory of experimental design. Chicago: University of Chicago Press, 1960; J. E. Freund, P. E. Livermore, and I. Miller, Manual of experimental statistics. Englewood Cliffs, N. J.: Prentice-Hall, 1960

The great classic is Sir Ronald Fisher's perennial, The design of experiments.[8]

[8] New York: Hafner Publishing Co. The first edition appeared in 1935.

Chapter 6. Cautionary Notes

As Stanley[1] has argued at length, psychometricians need to become much more aware that principles of experimental design should be kept fully in mind in the planning and execution of psychometric research. Coombs[2] showed rather long ago that the analysis of variance may be couched in correlational terms. We have extended the correlational possibilities for those who prefer to think in terms of mean r's. The crucial point, however, whether one prefers to report estimated variance components, $\bar{r}$'s, or both, is that experimental-design principles, particularly randomization within studies, apply to psychometric research but often are ignored there. It is especially dangerous to adopt the form of a statistical technique, such as the analysis of variance, without meeting its requirements.

Another, more specific precaution seems in order. Mean r's cannot attain values of -1 if more than two distributions are involved.[3] The smallest possible value of $\bar{r}$ among k paired arrays is as follows:

$$\min \bar{r} \geq \frac{-1}{k-1}, \text{ which is negative, and where } \bar{r} = \frac{\sum_{j=1}^{k} \sum_{h=1}^{j-1} r_{jh}}{(1/2)k(k-1)}, \; j \neq h.$$

---

[1] J. C. Stanley: Interaction of organisms with experimental variables as a key to the integration of organismic and variable-manipulating research, pp. 7-13 in 17th Yrbk Natl Coun. on Meas. in Educ., 1960; Studying status vs. manipulating variables, pp. 173-208 in Second Annual Phi Delta Kappa Symposium on Educational Research, 1961; Analysis of variance principles applied to the grading of essay tests, J. exp. Educ., 1962, 30, 279-283.

[2] C. H. Coombs, The role of correlation in analysis of variance, Psychometrika, 1948, 13, 233-243. Also see K. A. Brownlee, Statistical theory and methodology in science and engineering. New York: Wiley, 1960.

[3] This statement is proved in Appendix 2.

On the other hand, the highest possible upper limit, +1, is not a function of k.

Thus, if just two variables, such as height and weight, are correlated, an $\bar{r}$ of -1 is possible. If three are intercorrelated, an $\bar{r}$ as small as -0.5 is possible. For 101 variables, the minimum $\bar{r}$ is -0.01. This principle limits the negativeness of the various $\bar{r}$'s inferred from ANOVA mean squares.

Appendix 1

Expected values of mean squares under various mixed models and expected values of average variance and covariances under one model

Table 12

Expected values for mean squares for a four factor factorial design where I is random, and R, T, & S are fixed

$$E(MS_i) = \sigma^2 + RST\sigma_i^2$$

$$E(MS_r) = \sigma^2 + IST\sigma_r^2 + ST\sigma_{ir}^2$$

$$E(MS_t) = \sigma^2 + IRS\sigma_t^2 + RS\sigma_{it}^2$$

$$E(MS_s) = \sigma^2 + IRT\sigma_s^2 + RT\sigma_{is}^2$$

$$E(MS_{ir}) = \sigma^2 + ST\sigma_{ir}^2$$

$$E(MS_{it}) = \sigma^2 + RS\sigma_{it}^2$$

$$E(MS_{is}) = \sigma^2 + RT\sigma_{is}^2$$

$$E(MS_{rt}) = \sigma^2 + IS\sigma_{rt}^2 + S\sigma_{irt}^2$$

$$E(MS_{rs}) = \sigma^2 + IT\sigma_{rs}^2 + T\sigma_{irs}^2$$

$$E(MS_{ts}) = \sigma^2 + IR\sigma_{ts}^2 + R\sigma_{its}^2$$

$$E(MS_{irt}) = \sigma^2 + S\sigma_{irt}^2$$

$$E(MS_{irs}) = \sigma^2 + T\sigma_{irs}^2$$

$$E(MS_{its}) = \sigma^2 + R\sigma_{its}^2$$

$$E(MS_{rts}) = \sigma^2 + I\sigma_{rts}^2 + \sigma_{irts}^2$$

$$E(MS_{irts}) = \sigma^2 + \sigma_{irts}^2$$

Table 13

Expected values for mean squares for a four factor factorial design where R is random, and I, T, & S are fixed

$$E(MS_i) = \sigma^2 + RST\sigma_i^2 + ST\sigma_{ir}^2$$

$$E(MS_r) = \sigma^2 + IST\sigma_r^2$$

$$E(MS_t) = \sigma^2 + IRS\sigma_t^2 + IS\sigma_{rt}^2$$

$$E(MS_s) = \sigma^2 + IRT\sigma_s^2 + IT\sigma_{rs}^2$$

$$E(MS_{ir}) = \sigma^2 + ST\sigma_{ir}^2$$

$$E(MS_{it}) = \sigma^2 + RS\sigma_{it}^2 + S\sigma_{irt}^2$$

$$E(MS_{is}) = \sigma^2 + RT\sigma_{is}^2 + T\sigma_{irs}^2$$

$$E(MS_{rt}) = \sigma^2 + IS\sigma_{rt}^2$$

$$E(MS_{rs}) = \sigma^2 + IT\sigma_{rs}^2$$

$$E(MS_{ts}) = \sigma^2 + IR\sigma_{ts}^2 + I\sigma_{rts}^2$$

$$E(MS_{irt}) = \sigma^2 + S\sigma_{irt}^2$$

$$E(MS_{irs}) = \sigma^2 + T\sigma_{irs}^2$$

$$E(MS_{its}) = \sigma^2 + R\sigma_{its} + \sigma_{irts}^2$$

$$E(MS_{rts}) = \sigma^2 + I\sigma_{rts}^2$$

$$E(MS_{irts}) = \sigma^2 + \sigma_{irts}^2$$

Table 14

Expected values of mean squares for four factor analysis of variance, I& R random, T& S. fixed

$$E(MS_i) = \sigma^2 + RTS\sigma_i^2 + TS\sigma_{ir}^2$$

$$E(MS_r) = \sigma^2 + ITS\sigma_r^2 + TS\sigma_{ir}^2$$

$$E(MS_t) = \sigma^2 + IRS\sigma_t^2 + RS\sigma_{it}^2 + IS\sigma_{tr}^2 + S\sigma_{irt}^2$$

$$E(MS_s) = \sigma^2 + IRT\sigma_s^2 + RT\sigma_{is}^2 + IT\sigma_{rs}^2 + T\sigma_{irs}^2$$

$$E(MS_{ir}) = \sigma^2 + TS\sigma_{ir}^2$$

$$E(MS_{it}) = \sigma^2 + RS\sigma_{it}^2 + S\sigma_{irt}^2$$

$$E(MS_{is}) = \sigma^2 + RT\sigma_{is}^2 + T\sigma_{irs}^2$$

$$E(MS_{rt}) = \sigma^2 + IS\sigma_{rt}^2 + S\sigma_{irt}^2$$

$$E(MS_{rs}) = \sigma^2 + IT\sigma_{rs}^2 + T\sigma_{irs}^2$$

$$E(MS_{st}) = \sigma^2 + IR\sigma_{st}^2 + R\sigma_{ist}^2 + I\sigma_{rst}^2 + \sigma_{irst}^2$$

$$E(MS_{irt}) = \sigma^2 + S\sigma_{irt}^2$$

$$E(MS_{irs}) = \sigma^2 + T\sigma_{irs}^2$$

$$E(MS_{its}) = \sigma^2 + R\sigma_{its}^2 + \sigma_{irts}^2$$

$$E(MS_{rts}) = \sigma^2 + I\sigma_{rts}^2 + \sigma_{irts}^2$$

$$E(MS_{irts}) = \sigma^2 + \sigma_{irts}^2$$

Table 15

I and R random, T and S fixed

$$\overline{E(s^2_{rts})} = \sigma^2 + \sigma^2_i + \sigma^2_{ir} + \frac{T-1}{T}\sigma^2_{it} + \frac{T-1}{T}\sigma^2_{irt} + \frac{S-1}{S}\sigma^2_{is} + \frac{S-1}{S}\sigma^2_{irs} + \frac{(T-1)(S-1)}{TS}\sigma^2_{its} + \frac{(T-1)(S-1)}{TS}\sigma^2_{irts}$$

$$\overline{E(\mathrm{cov}(X_{rts}, X_{r'ts}))} = \sigma^2_i + \frac{T-1}{T}\sigma^2_{it} + \frac{S-1}{S}\sigma^2_{is} + \frac{(T-1)(S-1)}{TS}\sigma^2_{its}$$

$$\overline{E(\mathrm{cov}(X_{rts}, X_{rt's}))} = \sigma^2_i + \sigma^2_{ir} - \frac{1}{T}\sigma^2_{it} - \frac{1}{T}\sigma^2_{irt} + \frac{S-1}{S}\sigma^2_{is} + \frac{S-1}{S}\sigma^2_{irs} - \frac{S-1}{TS}\sigma^2_{its} - \frac{S-1}{TS}\sigma^2_{irts}$$

$$\overline{E(\mathrm{cov}(X_{rts}, X_{r't's}))} = \sigma^2_i + \frac{S-1}{S}\sigma^2_{is} - \frac{1}{T}\sigma^2_{it} - \frac{S-1}{ST}\sigma^2_{its}$$

$$\overline{E(\mathrm{cov}(X_{rts}, X_{rts'}))} = \sigma^2_i + \sigma^2_{ir} + \frac{T-1}{T}\sigma^2_{it} + \frac{T-1}{T}\sigma^2_{irt} - \frac{1}{S}\sigma^2_{irs} - \frac{1}{S}\sigma^2_{is} - \frac{T-1}{TS}\sigma^2_{its} - \frac{T-1}{TS}\sigma^2_{irts}$$

$$\overline{E(\mathrm{cov}(X_{rts}, X_{r'ts'}))} = \sigma^2_i + \frac{T-1}{T}\sigma^2_i - \frac{1}{S}\sigma^2_{is} - \frac{T-1}{TS}\sigma^2_{its}$$

$$\overline{E(\mathrm{cov}(X_{rts}, X_{rt's'}))} = \sigma^2_i + \sigma^2_{ir} - \frac{1}{T}\sigma^2_{it} - \frac{1}{T}\sigma^2_{irt} - \frac{1}{S}\sigma^2_{is} - \frac{1}{S}\sigma^2_{irs} + \frac{1}{TS}\sigma^2_{its} + \frac{1}{TS}\sigma^2_{irts}$$

$$\overline{E(\mathrm{cov}(X_{rts}, X_{r't's'}))} = \sigma^2_i - \frac{1}{T}\sigma^2_{it} - \frac{1}{S}\sigma^2_{is} + \frac{1}{TS}\sigma^2_{its}$$

Appendix 2

Theorem: $\bar{r} \geq -\frac{1}{k-1}$, where $\bar{r}$ is the average of the $(1/2)k(k-1)$ possible correlations among k columns of number taken two columns at a time.

Proof: Let z_{ij} be the column z-score for the ith row ($i = 1, 2, \ldots, n$) in the jth column ($j = 1, 2, \ldots, k$), i.e. $z_{ij} = \frac{x_{ij} - \bar{x}_{.j}}{s_j}$. This, then, implies that $\sum_{i=1}^{n} z_{ij} = 0$, and that $\sum_{i=1}^{n} z_{ij}^2 = n$.

In view of this, we note that:

$$\sum_{i=1}^{n}\left(\sum_{j=1}^{k} z_{ij}\right)^2 = \sum_{i=1}^{n}(z_{i1} + z_{i2} + \ldots + z_{ik})^2 = \sum_{i=1}^{n}(z_{i1}^2 + z_{i2}^2 + \ldots + z_{ik}^2 + z_{i1}z_{i2} + z_{i1}z_{i3} + \ldots + z_{i1}z_{ik} + \ldots + z_{i,k-1}z_{ik})$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{k} z_{ij}^2 + 2\sum_{i=1}^{n}\sum_{j=1}^{k}\sum_{h=1}^{j-1} z_{ij}z_{ih}$$

$$= \sum_{j=1}^{k} n + 2\sum_{j=1}^{k}\sum_{h=1}^{j-1}\left(\sum_{i=1}^{n} z_{ij}z_{ih}\right)$$

$$= kn + 2\sum_{j=1}^{k}\sum_{h=1}^{j-1}(n\, r_{jh}) = kn + 2n\sum_{j=1}^{k}\sum_{h=1}^{j-1} r_{jh}$$

$$= kn + 2n(1/2)k(k-1)\bar{r})$$

$$= kn + nk(k-1)\bar{r}.$$

Solving for $\bar{r}$, we find:

$nk(k-1)\bar{r} = -kn + \sum_{i=1}^{n}\left(\sum_{j=1}^{k} z_{ij}\right)^2$, which implies that

$\bar{r} = -\frac{1}{k-1} + \frac{\sum_{i=1}^{n}\left(\sum_{j=1}^{k} z_{ij}\right)^2}{nk(k-1)}$, which in turn implies that $\bar{r} \geq -\frac{1}{k-1}$,

with equality if and only if $\sum_{j=1}^{k} z_{ij} = 0$ for all i (i.e., if every z_{ij} equals -1 times the sum of the other $(k-1)$ z_{ij}'s in its row).